Sent as attachment

22 July 1958

JPRS/DC-241

ABSTRACTS
OF THE
CONFERENCE ON MACHINE TRANSLATION

(MAY 15-21, 1958)

PHOTOCOPIES OF THIS REPORT
MAY BE PURCHASED FROM THE
PHOTODUPLICATION SERVICE
LIBRARY OF CONGRESS
WASHINGTON 25, D. C.

U. S. JOINT PUBLICATIONS
RESEARCH SERVICE

Main Office:
Room 1125
205 E 42nd Street
New York 17, N. Y.

D. C. Office:
Second Floor
1636 Connecticut Ave., N.W.
Washington 9, D. C.

Ministry of Higher Education, USSR

First Moscow State Pedagogical Institute of Foreign Languages

ABSTRACTS

OF THE

CONFERENCE ON MACHINE TRANSLATION

(May 15-21, 1958)

MOSCOW, 1958

TABLE OF CONTENTS

PLENARY SESSION

THEORETICAL SECTION

SECTION ON ALGORITHMS OF MACHINE TRANSLATION

Page

Page

Page

PLENARY SESSION

1. THE METALANGUAGE OF MACHINE TRANSLATION AND ITS APPLICATION

N. D. Andreyer (Leningrad)

1. We call a metalanguage any linear system of signs used for the written designation of the elements in a particular system of ideas and the relations between these elements.

2. The class of metalanguages at the present time comprises mathematics, physics, chemistry, formal genetics, and symbolic logic.

3. The preparation of algorithms for machine translation requires the development of a special metalanguage in the symbols of which may be described the facts and relationships of the language systems that are subject to equivalent comparison.

4. The symbols used in the metalanguage of machine translation are regarded as metalanguage words and grouped in categories analogous to the parts of speech.

5. Types of commands in M.T. /machine translation/ are regarded as metamoods /META-NAKLONENIYA/

6. The use of metalanguage in the analytic part of algorithms.

7. The use of metalanguage in the transformational part of algorithms.

8. The use of metalanguage in the synthetic part of algorithms.

9. The possibility and value of a general theory of metalinguistic systems.

10. A comparative analysis of the class of metalanguages and the class of spoken languages may serve as a basis for elucidating the relations between formal logical semiotics and general linguistics.

2. SOME GENERAL PROBLEMS IN MACHINE TRANSLATION /M.T./

I. K. Bel'skaya (Moscow)

1. Experience gained in preparing experimental routines for machine translation from English, German, Chinese, Japanese, and Russian in the ITM and VT /Institut tochnoi mekhaniki i vychislitel'noi tekniki/Institute of Precision Mechanics and Computer Engineering/of the Academy of Sciences, USSR confirms the assumption that translation, even in such an unusual form as machine translation, is, as far as content is concerned, a linguistic problem.

- 1 -

2. The development of linguistic methods of solving M.T. problems may be achieved on the basis of so-called "traditional linguistics" and the results of such work may be of definite interest to linguistics.

The systematization of language phenomena that accompanies M.T. research should help to eliminate the well known contradictions and diffuseness in the definitions of certain linguistic categories accepted at the present time.

3. A distinction between the lexical and grammatical aspects of the translation problem seems essential. The difference in quality and degree of lexical and grammatical abstraction emerges in the system of machine translation with unusual clarity.

Rules of lexical character are recorded in a glossary. Grammatical rules are not included in the glossary and form the content of so-called "translation routines".

4. An M.T. glossary must be so constructed that its various parts can expand unevenly.

An M.T. glossary may be divided into 2 main sections:

    I   single-meaning glossary, and

    II multiple-meaning glossary.

Each of these is in turn subdivided into:

    Ia glossary of technical terms;

    Ib glossary of words in general use;

    IIa glossary of full-meaning words;

    IIb glossary of auxiliary words.

An M.T. glossary is accompanied by several auxiliary routines (comprising one cycle in the translation routine) in order that the lexical analysis of a sentence may be performed without human intervention:

    1. Routine of dividing a sentence into words /Routine 1 is not essential for all languages, only for such as Chinese, Japanese, Arabic, etc., where the sentence is written down in the form of an unbroken succession of signs with no spaces between the words/

    2. Routine of obtaining the glossary form of a word

    3. Grammatical analysis of "unknown words"

4. Syntactic analysis of "formulas"

5. Routine of distinguishing homonyms

6. Routine of analysis of polysemy.

5. The basic problems of an M.T. glossary—size and polysemy—are satisfactorily solved by combining the following two methods:

(a) division of the glossary into a series of "special glossaries" corresponding to various spheres of human activity (in our case - corresponding to the various branches of science);

(b) contextual (functional - semantic) analysis of the words.

6. The main features of an M.T. glossary are that it:

(a) contains a systematized description of each word that is capable of ensuring the subsequent grammatical analysis of the word in the sentence (the "invariant characteristics of the word");

(b) provides for a genuine correspondence between two lexical systems, registering the "relevant meanings" of words;

(c) takes cognizance of "zero meanings" of words, i.e. instances where a word must not be translated into another language as a separate lexical unit.

For the rest, an M.T. glossary may be arranged on the same principles as those underlying existing bilingual dictionaries. In particular, there is no need to convert an M.T. glossary into a "glossary of stems". Moreover, a glossary of words has definite advantages for M.T. too.

7. The solution of the problem of grammatical analysis in M.T. is connected with the realization of a logical, structural description of language. Hence, conclusions drawn from solving this problem may have a certain general linguistic interest.

8. Following the grammatical analysis of 5 linguistic systems—English, German, Chinese, Japanese, and Russian—for M.T., it seemed possible to use a consistent system of dividing words into the following 9 lexico-grammatical categories:

1. verbs,

2. substantives,

3. numerals,

- 3 -

4. adjectives,

5. adverbs,

6. prepositions /Chinese and Japanese postpositions may be classified as prepositions on the basis of their resemblance to prepositions in function/

7. conjunctions,

8. particles,

9. parenthetic words.

The principle of dividing words into these classes is similar to that underlying the division of words into parts of speech. Hence, there is no need to do away with the traditional names of the parts of speech. Only a bit more precision is required.

Thus, the classes of numerals, adjectives, and adverbs have been changed. Pronouns are not isolated in a separate class, but the pronominal category differs for such parts of speech as substantives, adjectives, and adverbs.

Systematization of grammatical categories within each part of speech resulted in differentiating between the variant (contextual) and invariant grammatical characteristics of the words.

9. The grammatical processing of sentences by the translation routines breaks down into two independent steps:

Analysis of sentence to be translated, and

Synthesis of translated sentence.

We call analysis routines that system of rules whereby the linguistic analysis of a sentence to be translated can be performed in such a way as to produce the information needed for the grammatical structure of the translated sentence.

In the M.T. variant developed at the Institute of Precision Mechanics and Computer Engineering of the Academy of Sciences, USSR, the analysis routines include the following 8 routines in cycle II:

1. functional analysis of punctuation marks;

2. breakdown of sentences into clauses and more precise definition of parenthetical phrases in clauses;

3. syntactic analysis of clauses;

4. "verb" routine;

5. "numeral" routine;

6. "substantive" routine;

7. "adjective" routine;

8. "changing word order in translated sentence" routine.

10. We call synthesis routines that system of rules whereby the grammatical structure of the translated clause can be formed.

As of now 4 synthesis routines for the Russian sentence have been worked out:

1. word-forming routine;

2. "verb" routine;

3. "adjective" routine;

4. "substantive" routine.

It is proposed to develop a routine for editing the style of translated Russian sentences as well as synthesis routines for several other languages, particularly Chinese and English.

This would make it possible to produce multilingual machine translation (from many languages into many languages), using Russian, it is suggested, as an intermediary language.

3. AN INTERMEDIARY LANGUAGE AND ARTIFICIAL INTERNATIONAL LANGUAGES

Ye. A. Bokarev (Moscow)

1. Creation of an intermediary language for machine translation or an artificial Esperanto-type international language requires the solution of several problems, the main one being the need to establish correspondences between the lexical and grammatical units of languages that differ in their structural characteristics.

2. International languages based on natural languages use everything that is essential for communication and reject what is non-essential or of little value (exceptions of various kinds, polytypic declensions and con-jugations, etc.). The most consistent in this respect are the autonomastic languages (Esperanto and Ido). Languages of another kind - the naturalistic (Interlingua and Occidental) - retain certain of the unjustified complications and inconsistencies of natural languages.

- 5 -

3. The most important problems in the field of grammar are: indication of the parts of speech, expression of subject-object relations, and word order in sentences.

4. In the field of word formation there is the problem of productivity of word-forming affixes and use of established patterns.

5. Some of these problems may be solved in various ways when an intermediary language or an artificial language for international relations is created. Nevertheless, there are many problems that can be solved in similar fashion.

4. THE VALUE OF MATHEMATICAL METHODS IN LINGUISTICS

R. L. Dobrushin (Moscow)

1. Uses of linguistics as a justification for its existence. Classical fields of use: teaching of languages and application to problems in history.

2. Demands on language research imposed by classical fields of application of linguistics.

3. Newest fields of application of linguistics: mechanical translation and use for transmission of information in the form of written and oral linguistic material.

4. Problems and methods of linguistic research dictated by the newest fields of linguistic applications.

5. Mathematical methods of linguistic investigation:

(a) methods used in theory of numbers applied to investigation of the grammatical structure of language;

(b) investigation of language structure by methods used in the theory of information;

(c) linguistic statistics.

6. Interrelations between classical and modern linguistic techniques. Potential for the development of mathematical methods.

5. CONVERSION OF COMMUNICATIONS AND CONVERSION OF CODES

V. V. Ivanov (Moscow)

1. In theoretical investigations dealing with automatization of linguistic processes, it is advisable to distinguish the conversion of communications (texts) from the conversion of codes (sign systems).

— 6 —

2. By communication conversion we understand the translation of a communication from one code into another (recoding) while retaining the invariant information. When speech is transmitted at a distance, the linguistic structure of the text is kept, which makes this case very simple. When sentences are converted within a single language the linguistic structure of the text is partially transformed. This transformation may, therefore, be regarded as a first approach to machine translation. In translating from one concrete language into another concrete language or into an intermediary language, it is possible to preserve the characteristics of the linguistic structure of the text, which are directly reflected on the structure of the text in the other language. In translating into the logical, abstract language of an information machine, only the logical structure of the text can be preserved. The increasing degree of difficulty of each of these tasks is determined by the complexity of the rules for converting a communication, which vary with the extent to which the information appearing as an invariant during the conversions can be formalized.

3. By code conversion we understand the translation of one code into another while retaining the code pattern. An intermediary language for machine translation and an abstract machine language for an information machine may be regarded as abstract systems, which are represented by the concrete language of scientific and technical texts. Therefore, to develop these abstract systems we require a formal analysis of the individual concrete languages in order to reveal their common patterns. An abstract machine language may be constructed by converting concrete languages derived, in turn, from interpreting an abstract language. The general theory of code conversion may be used for the deductive derivation of one scientific system from another. In this connection it is necessary to investigate code isomorphism in the various sciences (and code isomorphism in a single science at various stages in its history). At the same time a general theory of code conversion makes it possible to formulate with greater precision the concepts of comparative and historical linguistics due to the fact that comparative-historical calculation is a special case of code calculation.

6. THE SEQUENCE IN BUILDING A LANGUAGE SYSTEM

P. S. Kuznetsov (Moscow)

1. Any language is a system of simple units of various orders so interlinked by hierarchical relations that each elemental unit is in some respect indivisible (without loss of some of its properties) and at the same time consists of a certain number of units of a lower order.

2. The simple units of one order form what is called a level, stage, or layer in a language system. Thus, one level is formed by such elemental units as phonemes, another by morphemes, which consist of phonemes, a third by lexemes (words), which consist of morphemes, etc.

When we build any language system, apparently the simplest way should be to define in succession the units of the lowest order and then pass on to the units of the next higher order, the units and relations in which they must be defined in accordance with concepts already defined for the next lower order. Thus, having defined the concept of phoneme, we may define the morpheme, which always consists of a certain number of phonemes.

4. But if we proceed in this fashion, we shall not be able to construct an internally consistent system, since at certain stages along the way we will meet up with vicious circles (in the logical sense).

5. The reason is that a system of units in any single order requires certain concepts lying outside itself for its own construction or, in other words, forming with respect to it meta-concepts /META-PONYATIYA/. These meta-concepts relate in part to the system of units in a lower order (with respect to the order in question) and they may relate in part also to the system of units in a higher order (with respect to the order in question). Thus, the definition of phonemes and their interrelations (in the phonological sense, to which I subscribe; I have often set forth in print the case for this view, so there is no need for me to go into it again here) are based not only on concepts from the field of phonetics, but also on some concepts from the field of morphology, i.e., they relate to the level of morphemes.

6. A more complicated method of constructing a language system is outlined on the basis of the foregoing. In some cases it is necessary to proceed directly from the system of the lower (e.g. first) order not to the next higher (in the given case, second) order, but to the following (in our case, third) order; and having constructed it without utilizing the concepts of the second order, to proceed to this last; and then to return to the system of the third order and finish constructing it, now also making use of the concepts relating to the system of the second order.

7. <u>MACHINE TRANSLATION STUDIES IN THE MATHEMATICAL</u>
   <u>INSTITUTE OF THE ACADEMY OF SCIENCES, USSR</u>

A. A. Lyapunov and O. S. Kulagina (Moscow)

I. <u>Introduction</u>

1. Electronic computers are a highly efficient means of processing information.

2. It is practical to use electronic computers as an auxiliary tool for intellectual work.

3. Human speech as a means of transmitting information.

4. The importance of making it possible for machines to use human speech.

- 8 -

5. Machine translation as a first step in instructing machines to work with a language.

II. Brief Description of Work Done

6. French-Russian translation. Empirical formulation of rules. Construction of an algorithm suited to the machine's capabilities. Elaboration of problems connected with coding and information conversion in the machine memory and with the organization of programs to increase the efficiency of machine operation. Utilization of scales. Work on improving the algorithm and programs on the basis of experimental translations.

7. English-Russian translation. Use of structural-syntactic analysis of English. Classification of English and Russian words on the basis of formal criteria. Grammatical configurations of English and Russian, a comparison. Problems in eliminating homonomy. Use of experience with French-Russian translation in problems connected with coding, program construction, and Russian sentence analysis.

8. Problems in automatizing translation programming. Operational description of translation algorithms. Compiling program, constructing the translating program according to its operational description. Significance of experience gained in programming French-Russian translation.

9. Theory of numbers approach to the construction of a formal grammar. Classification of words, identification of configurations, determination of relations between words. Possibilities of using a similar approach to syntax and phonetics.

10. Basic principles of operation: advance by "ledges" /USTUPAMI7; maximal theoretical interpretation of each step; planning of work based on interrelations between machine and thought; close contact between groups working on different languages; joint work of mathematicians and linguists at all stages starting with the formulation of translation rules.

III. Problems

11. Linguistic problems in machine translation.

(a) Development of precise system of linguistic concepts, their operation in translation algorithms as a criterion of usefulness.

(b) Development of methods of constructing translation algorithms for different languages.

(c) Intermediary languages, construction and use.

(d) Problems in linguistic statistics.

(e) Investigation of language structure on the basis of translation algorithms.

12. Technical problems in machine translation.

    (a) Elaboration of effective designs for translation machines.

    (b) Establishment of operational systems for these machines.

    (c) Elaboration of special memory devices (large capacity with swift retrieval).

    (d) Design of special input and output devices.

13. Mathematical problems in machine translation.

    (a) Development of effective means of coding information at the various stages of operation.

    (b) Increasing the output of algorithms.

    (c) Investigation of abstract language models and translation models.

    (d) Elaboration of a mathematical language to describe translation algorithms.

    (e) Automatization of programming of translation algorithms.

14. Combined-cybernetic problems.

    (a) Machine output of algorithms.

    (b) Machine production of linguistic statistics.

    (c) Machine construction of models of concrete languages on the basis of limited text materials.

IV. Problems Connected with Work in the
Field of Machine Translation

15. Need to elaborate different approaches to the problem by different research groups maintaining close contact among themselves. Value of co-operation in machine translation. Need to establish systematic exchange of information between groups working in different cities.

16. Need for representatives of the various fields of specialization to participate in the work on machine translation: mathematicians, linguists, and engineers constantly cooperating at all stages of the work from formulation of rules to study of experimental translations.

8. AN INTERMEDIARY LANGUAGE MODEL FOR MACHINE TRANSLATION

I. A. Mel'chuk (Moscow)

The following represents one of the possible solutions to the problem
of machine translation from many languages into many languages:

    1. Two sets of rules are worked out for each language:

        (a) The rules of analysis which, with the help of appropriate
glossaries and charts, effect the transfer of a text into
a conventional numerical code in such a way that each word
in a given form and given syntactic function is matched
one-for-one with a chain of figures called set of infor-
mation for the word. The series of sets of information
developed is broken down into paired typical combinations
with which the relations existing in each given pair of
information sets have been matched one-for-one. The fixed
relation between the two sets of information (containing
the syntactic relation between the corresponding words) is
called a "configuration". One member of the pair which
satisfies the given configuration is called the "governing"
and the other the "governed" member. The total number of
configurations is not very large (in a specialized text--
no more than 200).

        As a result of the analysis, each word in the text
to be translated is replaced by a set of information and
each set contains an indication of what configuration it
satisfies and which member it is.

        (b) The rules of synthesis permit transition from the numerical
code, i.e., from the series of sets of information, to
words, to the actual text. This operation is the reverse
of analysis described above.

        Each configuration contains an indication of what
form a word that satisfies the configuration in question
as either member of the pair must have. Therefore, if
we know the stem of a word, the kind of configuration,
and exactly how the word satisfies it, we can synthesize
the necessary forms.

        Both analysis and synthesis are effected in com-
plete independence of the translation.

    2. A special system of rules and charts is being worked out:
determining correspondences between the conventional numerical code of
different languages (identical correspondences are not essential; rules for
choice may be used.) These correspondences are established on 3 levels:

(a) lexical correspondences (i.e. lexical transfer of stems);

(b) grammatical correspondences (transfer of so-called "extra-syntactic " categories as, for example, number in nouns or tense and mood in verbs);

(c) syntactic correspondences (correspondences between con-figurations - syntactic relations of different languages as well as correspondences between groups of configurations - clauses and various types of phrases).

This abstract system of correspondences is also called an intermediary language which does not exist, therefore, as any real or artificial language but represents a unique calculus.

3. The translation process consists of three steps:

analysis -- transition from a text in the source language to a series of configurations;

transition -- from a series of configurations in the source language to a series of configurations in the target language;

synthesis -- transition from a series of configurations in the target language to a genuine text in it.

4. Underlying the translation is a syntactic analysis: establish-ment of configurations, i.e., ascertaining the relations between words in the source language and expressing these relations by the most suitable means in the target language. Such morphological data as case, number, and person of a verb (also the use of auxiliary words is provisionally included here) are used only as aids while ascertaining the syntactic relations.

5. During the course of syntactic analysis both the functions of words in the sentence ("sentence members") and the interdependence of words are established. The latter factor is especially important, since the interdependence of words makes it possible during synthesis to regulate their arrangement, i.e. to achieve the best word order.

6. The model of an intermediary language that has been worked out for machine translation includes for the present Russian, English, Chinese, French, and Hungarian. The purpose is to develop a system of formulating rules and the best method of recording and arranging the material.

9. THE SIGNIFICANCE OF MACHINE TRANSLATION
FOR LINGUISTICS

M. I. Steblin-Kamenskii (Leningrad)

Besides promoting cooperation with representatives of the precise sciences and thereby instilling linguists with the need for greater accuracy in their research and formulations, work on machine translation is important for linguistics in three respects:

(1) It is critical of all the traditional grammatical concepts, primarily those like the "parts of speech", "members of a clause", "clause", etc. Based, as it is, on practical considerations, this criticism will be more objective and effective than purely theoretical criticism.

(2) It makes clear that the same linguistic fact may be described in various ways depending on what general definitions or terminological conventions are used, with the result that all the dogmas established in the individual branches of linguistics need to be reviewed.

(3) It will aid in overcoming linguistic "semantism" /SEMANTIZMA/, i.e. the practice whereby linguists follow the line of least resistance and study meanings, not the structure of language. Language differs from other sign systems not by the existence of meanings (which are not peculiar to language), but by the structure of expression.

10. THE "ACTIVE" AND "PASSIVE" GRAMMAR OF L. V. SHCHERBA
AND THE PROBLEMS OF MACHINE TRANSLATION

I. I. Revzin (Moscow)

1. The polysemantic term "grammar" (either "grammatical structure of a language" or "description of the grammatical structure of a language") is one cause of the erroneous conception that a given language has only a single grammatical structure, that there is only one correct "grammar" (as a description of a system).

2. The description of a language system depends on the goal that an investigator sets for himself. This notion was the core of the remarkable theory of L. V. Shcherba on "passive" and "active" grammar, which has suffered undeserved oblivion.

3. "Passive grammar studies the functions and meanings of structural elements in a language on the basis of their forms, i.e. the external side. Active grammar teaches the use of these forms." (L. V. Shcherba) The purpose of instruction in passive grammar is to teach one to understand a text in the language. The purpose of instruction in active grammar is to teach one to express thoughts in the language.

4. One of the dangers pointed out in connection with L. V. Shcherba's ideas is the assumption of a "denudation of thought" or "existence of thought without language" in passing from form to pure meaning and from pure meaning to form. However, no cognizance was taken of the fact that a thought need not be registered in a concrete language; it may be registered in an abstract, artificial language where there is a simple, reciprocal correspondence between the designator and the thing designated.

5. Machine translation assumes precisely such an abstract language, namely an intermediary language that must be implicitly present in any machine program and will apparently be described in the near future. If cybernetic analogies are adequately grounded, one may assume that the analogue of such an intermediary language is present in any translation (and, generally, in any form of logical activity).

6. Machine translation has demonstrated the correctness and need of a separate approach to the problem of text analysis ("passive" grammar in L. V. Shcherba's terminology) and to the problem of text synthesis ("active" grammar).

7. The first problem was effectively solved by purely formal means. The limits of machine translation depend on a full solution of the second problem (the compilation of a list of synonyms — by synonymy we understand the presence of several units corresponding to a single unit in an abstract language or what amounts to the same thing, a single unit of thought — and an algorithm for retrieving an equivalent under the given logical conditions).

8. Experience with machine translation has shown that, generally speaking, an inverse ratio is observable between the "active" and "passive" grammar of a language: the more complex the "passive" grammar, the simpler the "active", and vice versa. Hence, for a number of languages emphasis wholly on passive grammar might considerably alleviate the language curricula in schools.

9. L. V. Shcherba's ideas on the distinction between active and passive grammar, as strengthened and enriched by experience with machine translation, must ultimately find application in foreign language teaching (in secondary schools as well as in colleges and universities).

10. Secondary schools should make wide use of the methods of passive grammar, which are not only unusually effective for analyzing an unfamiliar text, but correspond to the habits of logical thinking developed in mathematics classes. Moreover, interest in learning the grammar of a foreign language can be heightened by introducing exercises in translating sentences "by machine". This would also serve the interests of polytechnical instruction.

— 14 —

11. The same considerations apply as well to language teaching in the natural science departments of universities and in the higher technical institutions where little use of the well developed formal-logical habits of students has been made up to now in foreign language teaching.

12. Creating a scientific theory of "active grammar" would not only push forward the frontiers of machine translation, but assist instruction in language schools where grammar is still taught in undifferentiated fashion. This is of particular concern to translation departments where necessity dictated the conversion of a theory of translation into a theory of active grammar.

11. A GENERAL THEORY OF TRANSLATION IN CONNECTION WITH MACHINE TRANSLATION

V. Yu. Rozentsveig and I. V. Revzin (Moscow)

1. The possibility of creating a scientific theory of translation is still being argued by a number of specialists, both linguists and literary critics. Nor has there been any final answer to the question of whether a theory of translation concerns scientific linguistics or belongs to the field of literature.

2. The polysemantic term "translation" also awaits a definition. The historical paramountcy of artistic translation has resulted in the conceiving of every translation as an artistic production, as a creative achievement in the realm of language. Meanwhile, the development of new types of translation activity, chiefly in the field of scientific and technical literature, has made another conception of translation urgent, i.e. as a process of establishing principles of correspondence between the structures of two languages.

3. Disclosure of the possibility of translating texts by a machine and development of a theory of machine translation has shown that distinguishing between the fields of translation makes limitation of both concepts logically inexorable:

(a) "translation$_1$" is translation as a form of creative activity and

(b) "translation$_2$" is translation as the establishment of strict correspondences.

Translation as a form of creative activity is an object of study for theorists of literature. Translation as the establishment of strict correspondences is an object of study for linguists.

5. A linguistic theory of translation must regard translation ("translation 2") as a special kind of decoding with subsequent encoding into another system of symbols. The distinctive feature of this transformation of information is in the irreversibility of the process. The reason is that simple, reciprocal correspondences between language systems are lacking. Hence, rules for correspondence in translation are complicated by the need to formulate a number of restrictive conditions. Determination of these conditions is a proper object for a linguistic theory of translation. A general linguistic theory of translation studies ideal types and routines for matching systems of language symbols; a particular theory of translation analyzes the correspondences between the two languages. A general theory of translation is chiefly a deductive discipline, while a particular theory of translation is inductive.

6. Thus, the methodology of a linguistic theory of translation comprises:

      (a) methods of structural comparative analysis or, in other words, analysis of the synchronous stages of various languages;

      (b) methods of linguistic statistics;

      (c) methods of logical semantics, more precisely general semeiology.

The very listing of these methods shows the main difference between the linguistic and literary theories of translation. The latter requires:

      (a) a study of the era;

      (b) world outlook and creative method of the writer and literary school;

      (c) peculiarities of his individual artistic style.

7. From the semantic point of view "translation $_2$" is a certain reflection in itself (a system of elemental meanings is assumed to be invariant). "Translation $_1$" from this point of view, is not a reflection in itself, since pragmatic meaning, which plays a major role in "translation $_1$", does not coincide in two languages.

8. Having marked off the object and methods of a linguistic theory of translation, we can not only ascertain the limits of machine translation, but also create a well structured, definitive theory of translation, that is to say a separate, scientific linguistic discipline. Creation of this discipline can help to perfect methods of training translators. It will undoubtedly find application in the teaching of foreign languages as well.

THEORETICAL SECTION

12. SPECTRA OF PHONEMES AND THEIR USE IN MACHINE TRANSLATION

V. A. Artemov and I. A. Zimnyaya (Moscow)

1. Oral information and translation machines must, among other things, be accessible to people with varying physical characteristics of speech. Therefore, their system of signalling must be based on the phonemic invariants of sounds or, in other words, on the spectra of phonemes.

2. Three aspects of the spectral analysis of speech sounds must be distinguished: (1) syntactic (phonologic), (2) semantic (phonetic), and (3) pragmatic (technical-communicative).

3. A syntactic investigation of spectra of phonemes is based on contrasts within the sound system of a given language. A semantic investigation relates the spectra of phonemes to word meanings and grammatical forms. A pragmatic investigation of the spectra of speech sounds originates in and services practical needs.

4. A syntactic and semantic investigation of spectra of phonemes provides an exhaustive analysis of their physical properties which form structures bearing a comparative and systematic character.

5. A pragmatic investigation of spectra of phonemes requires the determination of their minimal characteristics, which permit of their full or partial restoration, i.e. it becomes a compression of the spectra of phonemes. A pragmatic investigation of spectra of phonemes becomes their compandor, including the compression and expansion of amplitude.

6. The Laboratory of Experimental Phonetics and Speech Psychology (LEF and PR) /Laboratoriya eksperimental'noi fonetiki i psikhologii rechi/ of the First Moscow State Pedagogical Institute of Foreign Languages (MGPIIYa)/Moskovskii gosudarstvennyi pedagogicheskii institut inostrannykh yazykov/conducted investigations of the spectra of 5 vocalic phonemes of a, o, u, i, e type in the following languages: (1) Russian (V. A. Artemov and I. A. Zimnyaya), (2) Georgian (T. G. Tsibadze), (3) Armenian (A. M. Aramyan and A. A. Khachatryan), (4) Lettish (I. A. Zimnyaya), (5) Albanian (I. A. Zimnyaya), (6) Bulgarian (I. A. Zimnyaya), (7) Czech (I. A. Zimnyaya), (8) German (L. P. Blokhina and I. A. Zimnyaya), (9) French (K. K. Barashnikova and V. S. Sokolova), (10) English (I. A. Zimnyaya). In addition, data on English were drawn from the works of Paget, Green and Potter, Petterson, and Kopp for purposes of comparison with the studies of the LEF and PR.

7. All the material was recorded with a basic tone of 120-150 cycles per second at a level of 65-70 db. The pronunciation of each speaker was representative of the literary speech of the various languages.

8. A comparison of the guantitative and graphic data shows that the following pragmatic rules are observable within each language:

(a) the a-type vowel is characterized by a wide formant region (600-1200 cycles) with gradually increasing intensity of the components in the direction of high frequencies (1200-2500 cycles).

(b) The o-type vowel is characterized by a central formant region somewhat shifted down to 400-1000 cycles per second.

(c) The u-type vowel is characterized by a somewhat narrower central formant region shifted still further toward the low frequencies of 300-800 cycles per second with a maximal elevation of amplitude in the range of 300-350 cycles per second.

(d) The i-type vowel is characterized by two main formant regions. The first is in the range of lower frequencies and almost coincides with the range of maximal intensification in the main formant of the u-type vowel, as Paget has pointed out. But a gentle falling-off is observed in amplitude of the u-type, and a steep falling-off in the i-type.

(e) The e-type vowel is distinguished from the i-type by the formants shifted more to the center. The broader the e, the closer the formants come together.

9. The above-mentioned acoustical properties of the vowels completely correspond to the position and operation of the resonance chambers of the vocal apparatus, as stated in several reports of the LEF and PR as well as by Paget and Yakobson.

10. These studies indicate that the spectra of vowels on the syntactic and semantic plane have a structural character. V. A. Artemov suggested a means of determining these structures. It consists of separating from the vowel spectrum all the areas of relative intensification and establishing correlations between them, taking the lowest of them as 1.

11. At the same time a comparison of the spectra of the 5 types of vowels studied indicates that a structural correlation between the areas of intensification is retained within definite limits in the languages investigated. In this connection it is possible to speak about a certain structural and comparative invariant of these types of vowel spectra in the various languages, which is essential for signalling technique in translation machines.

13. AN OBJECTIVE INVESTIGATION OF MEANING ASSOCIATIONS

O. S. Vinogradova and A. R. Luriya (Moscow)

1. An objective investigation of the association of meanings that are aroused in man by some word or other is a basic necessity for psychology as well as for linguistics.

Despite the considerable progress achieved by modern linguistics, information theory, and psychological investigation of the development of the meaning of words in children, objective research techniques both of potential associations aroused by words and of the dynamics of these associations still remain to be worked out.

2. The use of different variations of the conditioned reflex method may play a vital role in elaborating objective ways of investigating meaning associations. By combining the showing of a word with some kind of involuntary reflex responce (vascular, cutaneous-galvanic, etc. reaction) and then showing other words, the investigator is in a position to establish objectively what group of words shown elicits similar reactions and is consequently, to some extent, the equivalent of a previously shown word; and at the same time he is in a position to trace both the structure and the dynamics of these associations.

3. The report discusses the results obtained from an objective investigation of the system of meaning associations by registering the specific and non-specific conditions of vascular reactions. Conclusions are drawn concerning certain factors that may determine the structure and dynamics of these associations in normal and abnormal experimental subjects.

14. THE TREATMENT OF CERTAIN CONCEPTS IN STRUCTURALISM

V. I. Grigor'yev (Moscow)

1. Interest has grown of late in the methods and concepts of the structuralist approach in linguistics due to the development of machine translation and other branches of applied linguistics. However, recent articles have treated certain structuralist concepts in an excessively one-sided manner and, in essence, incorrectly.

2. Phonemes are treated as though they were connecting elements lacking physical reality. The physical character of the differential signs of phonemes is denied. Real speech sounds are represented as something external with respect to language. Meanings, which are also removed from language, receive the same treatment. This method of handling speech sounds and meanings reflects only the views of L. Yelm'slev's group and is not to be ascribed to structuralism in general.

3. Actually, the structuralist method of investigating speech sounds takes into account their acoustic and articulative properties. The functional criterion used by the structuralists in phonetics makes it possible to isolate from the entire diverse mass of phonetic material the physical (acoustic and articulative) properties that carry the functional load and, consequently, are of prime importance to the linguist. The functional criterion ensures a differentiated (from the viewpoint of language structure) approach to the varied and changing properties of phonetic material. Using the functional criterion, linguists may be very helpful to engineers in solving practical problems confronting the several branches of engineering; contrariwise, orientation on pure relationship elements would prevent the linguists from solving practical problems and do away with the possibility of cooperation between them and the engineers.

4. The attitude of the structuralists toward meaning was determined by their interest in working out an objective method of investigating language. The striving to escape from the inadequacies of traditional linguistics led the structuralists to refuse in general to consider meaning as a solid criterion of linguistic form. However, this refusal to take account of meaning in research methodology does not determine the structuralists' theoretical treatment of meaning. In many cases it exists harmoniously side by side with the acknowledgment of meaning as a basic element in the functioning of language. It must be admitted, however, that rejection of the semantic criterion imposed severe limitations on this school of linguistics. In practice, the field of semantics remained outside structural analysis.

5. The meaning of a word is the linguistic form of expressing an idea. Meaning cannot be separated from language simply because it does not exist prior to or apart from language. At the same time, meaning is a basic factor of language, determining its structure. It is important for the further development of applied linguistics that objective methods of semantic analysis be worked out. Naturally, in solving this problem full use will have to be made of the experience gained by the representatives of structuralism in their objective investigations of language.

6. A critical exploitation of the experience of structuralism is scientifically advisable. It is an indispensable preliminary stage in the task of introducing mathematical research methods into linguistics.

15. THE SIGNIFICANCE OF FREQUENCY AS A FACTOR IN DETERMINING THE STYLISTIC FUNCTIONS OF WORDS

V. M. Grigoryan (Yerevan)

1. A comparative study of modern Russian-Russian dictionaries reveals contradictory data. Thus, in various dictionaries (e.g., Monotypic 4-volume works) one and the same word may be defined in different ways from the

viewpoint of the language's limitations with respect to stylistic usage; and it is often possible to find inconsistencies in the order in which meanings are arranged. These (and other) contradictions make things difficult for the reader who seeks information in order to determine the operative norms for a given linguistic fact.

2. Since the norms, as a rule, are correlated with the factor of frequency, statistical data are extremely essential in many cases, if maximal precision is to be attained. Some considerations supported by Russian language data (with due regard for strict synchronousness) are cited by way of illustrating this contention.

3. The plan proposed by us is not original; it agrees in principle with that employed in several frequency dictionaries published abroad (Harry H. Josselson. The Russian word count, Detroit, 1953; Victor Garcia Hoz. Vocabulario usual vocabulario comun y vocabulario fundamental, Madrid, 1953.).

They are usually constructed on the basis of the familiar correlation between style and genre. Adopting this plan on the whole, we propose to set up 4 categories: (1) verse (2) speech in dialogues (3) speech in monologues - using material from fiction (4) non-fiction literature - newspapers, documents, etc. It is obvious that statistical data reflecting the frequency of usage of a specific word in each of the 4 categories must be selected on the basis of equal conditions. Clearly, these equal conditions will be ensured if the frequency of a given word is derived from an equal number of words in all 4 categories. If we designate the categories by a, b, c, and d, respectively, the total preliminary number of words in category a must be equal to the total preliminary number of words in category b, etc. This word total, it seems to us, can be advantageously determined by using the Hoz method. In addition, selections must be made from purely random material (but within the given category); the more varied the material, the more accurate will be the information.

The resultant data can be used to determine stylistic functions.

16. AN EXPERIMENT TO DEFINE THE CONCEPT OF GRAMMATICAL CATEGORY

R. L. Dobrushin (Moscow)

A given finite number of words is examined. A finite, ordered aggregate of words is called a sentence. The division of all sentences into two non-crossing classes is assumed to be given: a class of grammatically valid sentences and a class of grammatically invalid sentences. Word A is called subordinate to word b, if a valid sentence containing word A remains valid after A is replaced by b. Two words A and b are called equivalents, if A is subordinate to b and b is subordinate to A. All words are divided

into two non-crossing classes of words equivalent to one another. Class A is subordinate to class B, if all words entered into class A are sub-ordinate to words entered into class B. The system of classes and sub-ordinations thus obtained - called the basic grammatical structure of the language - is examined. The result is a definition of the concept of grammatical category.

17. THE THEORY OF PROBABILITY AND DETERMINATION OF LINGUISTIC RELATIONSHIP

A. B. Dolgopol'skii (Moscow)

The proposed method of determining the relationship of language families by applying the theory of probability is, in broad outline, as follows:

1. On the basis of linguistic experience, those semantic points are isolated in which maximum historical stability of morphemes (without borrowing) is observed.

2. A determination is made in each group of languages under con-sideration as to which morphemes possessing a given meaning may with greater probability be regarded as the older. The usual techniques of comparative historical research as well as the method of internal reconstruction are used for this purpose.

3. We cannot speak about phonetic correspondences between language families being compared before the fact of relationship has been established. Hence, at this initial stage of investigation we must rely wholly on phonetic resemblances. More precisely, we rely here on subsequent probability cor-relations. Following a comparison of cognate languages, it appears that the n-sound is the most probable of all the sounds in any single related language that correspond etymologically to the n-sound of another related language. The same may be said of the m-sound. But, possibly, not of the s-sound. At any rate, among all the sounds that correspond in one language to the s, z-sounds in another related language, the most probable, apparently, are the sounds of the same s, z-group. This would also seem to be true of the l, r-group, the b, p, f-group, the t, d-group, the k, g, k, h-group, etc. In this connection, we perhaps can't say anything about vowels or laryngeals. Starting with these probability considerations, we may be able (leaving aside the vowels and laryngeals) to base our subsequent discussions on the data of consonant coincidences between various morphemes in the different languages. We will term "consonant coincidence" the correspondence between consonants that remain within one of the above mentioned groups. These groups must be chosen in such a way that phonetic shifts of these sounds are no more probable than retention of the sound (retention within the group). The groups cited here are obviously only for illustrative purposes. Actually, comparative-historical phonetic materials from all possible language families must be used to establish the most probable sound correspondences (one of

these correspondences is the most probable for several sounds - the cor-
respondence of a sound to itself). As a result, we may select, let us say,
10 or 7 different sound units which will constitute the material for
phonetic comparisons.

    4. Comparing the equivalent morphemes in the different families,
we note the phonetic coincidences (of para. 3). We then use appropriate
formulas from the theory of probability to measure the probability of the
accidental coincidence of a certain number of morphemes in so many languages,
from so many comparable items, taking into account the number of old
synonymous morphemes for each semantic point of each language group as well
as the total number of consonants distinguished during the comparison.
(Cf. para. 3).

    If the probability of accidental coincidence proves to be quite low,
it will be a weighty argument in favor of the relationship of the languages
in question.

    Use of the theory of probability will enable us to test the evidence
from comparisons between the various languages cited in numerous works
dealing with the problems of language family relationship (e.g. Trombetti,
Winkler, etc.).

18. A GENERAL THEORY OF DEFINITION AND THE POSSIBILITY OF APPLYING IT TO THE THEORY OF TRANSLATION DEVICES

A. A. Zinov'yev (Moscow)

    1. The process of translating from one language into another may be
described as a language consisting exclusively of definitions. Breakdown
of the language into elements is here assumed to be effected. It is
possible to model the formal apparatus of definitions, one may suppose,
by means of a special device. Having determined all possible definition
type relations at least between a selected part of the elements of one
language and a selected part of the elements of another language, we can
use the modelling device to produce in standard form at least partial
translation (if only in initial approximation).

    2. A general theory of definitions is constructed as part of a
theory of symbols. Several variants are possible depending on the original
concepts in the statement and on the formal apparatus for constructing the
theory. The suggested variant is characterized by an initial concept
"Choice," a special means of defining the concept "Symbol", "Term", and
"Definition". The formal apparatus is constructed on the basis of the
functors ⟨⟨·⟩⟩ ("Each of") and ⟨⟨ : ⟩⟩ ("anyone and only one of") and on the ad-
mission as an initial logical connection of such connection as could possibly,
with some limitations, be represented as a formal implication without con-
traposition.

- 23 -

3. A general theory of definitions may contain proof of rules for definitions, elicitation of the conditions governing their use, rules for deduction and their interconnections.

19. LINGUISTIC PROBLEMS CONNECTED WITH POETRY TRANSLATION

V. V. Ivanov (Moscow)

1. The distinction between the poetic model of a text and this text may serve as a convenient starting point in solving the problem of poetry translation. Translation makes it possible to recreate the same poetic model by means of another language while retaining the relation between the model and the text. On the other hand, the direct conversion of a poetic text in one language into a poetic text in another is impossible.

2. The amount of information contained in a text is determined by the extent of deviation of this text from the statistical norms of ordinary language and from the statistical norms of the poetic language of a given era. A violation of the statistical norms of ordinary language may become the norm of poetic language, which results in decreasing the amount of information contained in poetic texts. Poetry translation assumes the transmission of the statistical characteristics of a text in conformity with the language into which the translation is made.

3. The sound structure of verse is determined by the phonological structure of the language, as was first pointed out by R. Yakobson. It follows from this that transmission of the phonetic characteristics of the text structure is possible only when the corresponding elements in the phonological systems of the two languages coincide. The non-translatability of a poetic text is to a very large degree determined by the fact that in poetic language the plane of content is functionally connected with the plane of expression; insofar as the plane of expression is in principle un-translatable, the plane of content appears partially untranslatable. This limitation may also apply to the poetic model of the text, if (as with Khlebnikov) units from the plane of expression are included in this model.

4. Phonetic coincidences of parts of words are used to organize a poetic text chiefly in cases where they are superfluous from the morpholog-ical point of view. Conversely: morphologically essential phonetic re-semblances contain the least amount of information from the viewpoint of poetic organization of speech (cf. the problem of verbal rhythm in Russian poetry). Consequently, the possibility of transmitting phonetic repeats /POVTOROV/ depends not only on the phonological, but also on the morphological resemblances between the two languages in question.

5. The predominance of syntagmatic connections between words over paradigmatic connections is a peculiarity of poetic text on the plane of content. We may see in this the result of transforming language text in

accordance with a poetic model (the relation between "emotion and text"
/PORYVA I TEKSTA/, to use O. Mandel'shtam's terms). This transformation
can be effected not only in the original but also in the translation.

6. A line-by-line analysis of metres does not yield satisfactory re-
sults because it reveals little about the rhythmic structure of long
passages which are the real units of poetic speech (cf. the definition of
a period as "wave length" in Milton's verse, as suggested by T. S. Eliot).
If a line-by-line translation appears impossible, then for a translation
based on a poetic model of the entire text it is considerably more im-
portant to translate the major rhythmic and syntactic units into which
the work is divided; as an example, Rilke's translation of VYKHOZHU ODIN
YA NA DOROGU /I go out alone onto the road/is analyzed. The continuity of
an invariant text model and its inability, in principal, to be formalized
/NEFORMALIZUYEMOST'/exclude the possibility of automatic translation of
poetry by modern computers.

20. HEGEL'S THEOREM AND LINGUISTIC PARADOXES

V. V. Ivanov (Moscow)

1. The resemblance between mathematics and linguistics also applies
to the trends of these sciences as they develop in the 20th century. The
theoretical foundations of the sciences are being investigated in anticipation
of practical applications; the results of these investigations will eventually
prove vital for practical purposes.

2. Hegel's theorem, according to which the non-contradictableness of
a theory cannot be demonstrated within the formalized theory itself, may be
extended to linguistic theories by means of Lotze's generalization of the
theorem, which comes down to an affirmation of the incompleteness of any
system of symbols (including language). However, it would be essential
not to restrict oneself to this formulation in investigating the founda-
tions of structural linguistics, but to examine the conclusions resulting
from a linguistic analogue of Hegel's theorem.

3. The most severely formalized theories of language that examine
constructional linguistic objects were developed within the framework of
distributive analysis, which assumes the possibility of describing the
elements of a language on the basis of their distribution. It is not
difficult to show that the logical application of this principle leads to
linguistic paradoxes (e.g. in the distributive separation of phonemes,
word classes, meanings of polysemantic words, etc.). The distribution of
elements turns out to be impossible, if these elements were not given
previously. But the axiomatic introduction of language elements contradicts
not only the principles of distributive investigation, but also the re-
quirements of automatic analysis of written and oral speech. The axiomatic
introduction of a class of regular sentences appears to be unsatisfactory
for purposes of synthesis.

4. For the reasons given above, it is possible at the present time to fashion a formal theory, which can be used to construct a program of automatic analysis, only for a maximally simplified approximation to a real language. We have in mind cases with simple correspondence between form and substance: on the plane of expression - for a system of standard, typical variants of phonemes, on the plane of content - for a standard language of science. The absence of paradoxes when these cases are analyzed does not permit, however, of extending the results obtained to ordinary language, the metalanguage for which (unlike the cases mentioned) cannot be formalized (this applies both to the phonological and to the semantic metalanguage). Automatic analysis of real language requires the employment of linquistic methods other than those considered above and the use of self-teaching type machines (with probability elements).

21. METHODS OF BREAKING DOWN A SYNTACTIC WHOLE

L. I. Iliya (Moscow)

1. Linguists representing the most different schools use as a starting point in their methods of analysis the possibility -- objectively existing in any language -- of isolating a certain "whole" as a maximal unit that can be broken down into similar segments, i.e. comparable in any respect whatsoever. This "whole", which has been variously called "utterance", "sentence", or "clause", belongs simultaneously to all the planes or "levels" of a language -- phonological, grammatical, and semantic -- and is character- ized by the fact that its borders coincide in all three planes, which makes this segment a maximally complete or basic unit for any decomposition.

2. The breakdown of a "whole", due to its complexity of structure, is done on the basis of criteria that differ for each plane. As a result, it yields segments the boundaries of which do not always coincide or, as they say, are not "commensurable". Semantic decomposition is to a certain extent independent of the grammatical, and it fails to establish a fixed correlation between the boundaries created by rhythmic-intonation de- composition and the boundaries of morphemes, words, and groups of words.

3. Modern linguists have attempted to eliminate the incommensurability of the planes by seeking a single principle common to all stages of analysis. However, unity of principle is achieved in some theories by ignoring some aspect of language structure (e.g., meaning is excluded in Harris' method and in rhythmic-intonation decomposition of Trager and Smith, while gram- matical structure is ignored in Shcherba's intonation-semantic decomposition). Orderliness of method is attained at the price of simplifying linquistic analysis, which therefore cannot be regarded as adequate for research in all its complexities. However, new methods of analysis focussing on for- mal criteria have been used to study them deeply, and modern techniques of measuring such language units as phonemes, morphemes, and words have reached a high degree of precision.

4. The task of linguistic analysis is not only to isolate the basic linguistic units, but to determine the relations between the units that allow of semantic relations. The contemporary school of linguistics acknowledges as "structural," i.e. which deal with linguistic analysis, only those relations to which definite forms of expression, "signals", correspond.

Two main trends in the investigation of syntactic relations can be discerned at the present time: (a) the comparatively recent theory of "direct constituents" (Bloomfield, Pike, Wells), which bases sentence decomposition on the relations of a logical hierarchy of subordination that links all the parts of a sentence into a single whole, and (b) the theory, which may be provisionally called the theory of "members of the sentence". It has a long tradition and many opponents, but finds support among the major representatives of contemporary linguistics (Kurilovich, Bazel in part, Diederichsen). The theory considers the sentence a whole, the parts of which are linked together by functional relations.

5. The direct constituent method, which is based on a single type of relationship—the heterogeneity of functions of the constituents—leaves the general problem of determination of syntactic relations open and investigates for the most part the combinability of constituents and typical patterns.

On the other hand, for the theory of "members of the sentence" the problem of syntactic relations is fundamental. Formerly, these relations were all too frequently distinguished purely on the basis of meaning, not of formal criteria, although the inclusion of such criteria in the principle is desirable and feasible (Friese, Togebyu). The study of basic syntactic relations requires for its own continuing development that all modern methods of linguistic analysis be utilized, particularly the technique of distributive analysis.

The "direct constituents" and "members of the sentence" methods do not exclude one another. Rather, they are complementary, as they permit the sentence to be studied in various respects.

22. THE LOGICAL NATURE OF CONTEXT

G. V. Kolshanskii (Moscow)

1. The term "context", given the polysemia of language forms, may be defined from the linguistic point of view as a combination of conditions determining the simple, concrete identification of any linguistic phenomenon (lexical, grammatical, etc.). By "simple" we are to understand the display of only one of the many possible properties of the form under the given conditions (e.g. one meaning of a word, one word order, one intonation, etc.). In this report we are considering cases of determination of meaning in polysemantic words regardless of the method of origination (metaphor, metonomy, homonymy, etc.).

2. Contextual conditions may be found within the language itself, but they may also comprise indications lying outside language material. Among the language conditions it is necessary to distinguish between indications included within a single sentence and textual indications. Among the external conditions it is necessary to differentiate between situation, object, and graphic indications.

3. The combination of possible conditions called context may be realized while the precise meaning of a sentence is being formulated in language only through a definite, active, logical process, since indications by themselves are inert and can influence the meaning of a linguistic form only as a starting point in the functional process of achieving a result that makes sense.

Since the method of search by context is effective in the semantic area of language, it is in essence a speculatative, logical process of reasoning about the meanings of language forms.

This rational search for the essential and uniquely correct (in the ideal approach to a solution of this problem) result is a process of constructing a syllogism or chains of syllogisms where the answer needed to establish the true meaning of the word and sentence is the final deduction.

4. A syllogism is constructed by searching for the appropriate premise of a universal hypothetical syllogism (if.....then) or by interpreting a hypothetical-disjunctive judgment, the complexity of which depends on the character of the indication underlying the premise.

While searching for the unknown meaning through external extra-linguistic indications, a syllogism is formed in accordance with the nearest indication contained in the context (e.g. determination of the meaning "table" as a piece of furniture in the sentence"He has a good table" is based on situation. The meaning "table" may be "either A or B". Here it is not B. Therefore it is A, i.e. a piece of furniture.

5. Mention of the subject is sufficient for the major premise in order to determine the meaning through objective context (e.g. determination of the meaning of the word "solution" in chemistry and electrical engineering is made in similar fashion). The form of writing in a written text may also serve as a starting point for a syllogism about the true meaning of a word (e.g. a foreign spelling).

6. The method of searching for the determining factor through lexical environment is the most familiar way of determining the meaning of linguistic forms. The premise is based on the immediately adjacent word (starting of a Sputnik, starting of a motor) and a word standing in any position in the appropriate group (an effective operation to destroy.....a hostile garrison, vermin, tumors, etc., where all the semantic variants

are included as members in the major premise of an hypothetical-disjunctive syllogism).

7. If there are insufficient resources within the sentence, the true meaning of a word is sought by forming several syllogisms to search for the premise of the last conclusion on the basis of the entire paragraph or text (e.g. "we did not allow our house to be harmed" receives the following logical interpretation:

If it is not a question here of a concrete house and family, then the word "house" must be understood to mean "company" /=business/. After examination of the text, the first two senses are set aside and the meaning "company" remains as an affirmative answer according to the rule for a disjunctive syllogism. In German alle Raeder stehen still /all the wheels are standing still/--a similar analysis for the meaning--all traffic came to a stop).

8. The process of ascertaining the true meaning may be logically carried out by a hypothetical-disjunctive syllogism, but depending on the nature of the desired result the conclusion may be reached either by eliminating parts of the disjunctive judgment (given the possibility of complete enumeration of all the meanings of a word) or by first forming a disjunctive judgment (meaning, the word A may be either, or...). It should also be kept in mind that each operation is subject to rechecking.

9. Due to the fact that analysis of context is essentially a rational, logical process, it can in principle be theoretically formulated as an ordinary logical operation and be performed by a machine. The feasibility and advisability of any arrangement in connection with machine translation is the decisive factor in a given case.

For simple formulas in a context the formalized operation to search for the necessary meaning may be worked out by introducing a simple quantor (a thematic quantor). When the meaning of a word is being interpreted on the basis of immediate environment, a virtually disjunctive premise may be set up, obviously consisting of up to 3-5 words occurring before and after a polysemantic word, provided, however, that preliminary linguistic analysis determines all the cases where the meaning of the given word depends on words capable of being associated with it. At the present stage this work can be performed only for a limited group of words in certain texts.

10. If the context extends beyond a sentence, the solution of a disjunctive syllogism is practically impossible, since one cannot formally mention the indications on the basis of which the parts even of a fully set up disjunctive judgment will be eliminated.

Likewise formally insoluble is the problem of ascertaining the limits of the operation to analyze context (both within a sentence and, much more so, outside it). This is the realm of active, creative thought. Thus, a complete, practical solution to the problem of mechanical determination of word meaning by context is excluded. Formalization of the rules for interpretation of context in machine translation clearly requires the application of statistical methods for probability determination of the contextual meaning of the words.

23. LINGUISTIC STATISTICS FROM RUSSIAN TEXTS

R. G. Kotov (Moscow)

1. The development of machine translation and the application of methods of analysis and syntheses to communications technology have created sound conditions for expanding cooperation between linguists and engineers. In this connection there has arisen a need to introduce into linguistics objective research methods permitting mathematical handling of the data. Linguistic statistics, which operates with quantitative values, offers wide possibilities for linguistic research. Linguistic statistical data are used to solve a number of problems in machine translations and communications technology. In addition, they may be successfully exploited for lexico-graphical purposes and for foreign language teaching.

2. The current statistical investigation of Russian language texts aims at preparing preliminary data in connection with constructing the program of lexical coding of telegraphic messages. The work was first done by hand on specimens of texts containing a total of 20,000 words. Methods of analysis were determined by the existing possibilities and research goals. The texts to be analyzed were entered in order on index cards in the form of two-member word combinations, which made various types of calculation possible.

It is proposed to use in the future machine methods for several tabulations, e.g. word frequency.

3. The treatment of the material has yielded thus far a frequency glossary, glossary of stable word combinations, and data on the frequency of endings. Some principles governing the statistical distribution of the glossary for the texts examined are elucidated on this basis.

4. Superfluousness in Russian texts of the type investigated is being determined by taking cognizance of probability correlations in the glossary. A theoretical limit to the savings expected from lexical coding is being ascertained. Lexical coding is regarded here as a particular case of de-correlation /DEKORRELATSII/messages by consolidation /UKRUPNEN'YA/.

5. Work is going on to : Elucidate the main types of two-member word combinations the sequence of which makes up a text, ascertain the provisional probabilities of endings, and eliminate the uncertainty of choice of grammatical form in relation to the preceding word. Data obtained on the material of two-member word combinations are assumed to apply to multi-member word combinations and to the sentence as a whole.

24. A METHOD OF DEFINING GRAMMATICAL CONCEPTS

O. S. Kulagina (Moscow)

1. Inconvenience of existing grammatical systems for machine translation and need to elaborate precise definitions of concepts.

2. Initial base of undefined concepts: word sentence and OTMECHENNAYA sentence, environment.

3. Breakdown of multitudes of words into submultitudes /PODMNOZHESTVA7, consolidation of breakdowns.

4. Concept of B-equivalence, amalgamation of B-equivalent submultitudes. Derived breakdown. Theorem concerning the impossibility of secondary amalgamation by equivalence.

5. Sequence of amalgamation of words: families, classes, types. Concept of a simple language. Two definitions of type and their equivalence.

6. Determination of configuration, resultant element, ranks of configurations.

Concept of subordination of configurations.

Determination of relations between elements of configurations.

25. A FORMAL THEORY OF THE SENTENCE

I. I. Revzin (Moscow)

1. More than 200 different definitions of "sentence" make, on the one hand, a deductive development of syntax impossible and show, on the other, that the approach to the problem of defining basic linguistic units requires greater precision.

2. Any definition of a language element is a metalinguistic expression (explicit or implicit). "Sentence" as a language word is, by its nature, different from "sentence" as a metalanguage word. Therefore, the aim to include in the definition everything that is intuitively understood when the sound complex "sentence"is pronounced is scarcely realizable. A term in linguistics, like an expression in metalinguistics, may reflect only

certain intrinsic features corresponding to the usage of the word.

3. An analysis of existing definitions of "sentence" makes it possible to divide them into two unequal groups. The overwhelming number of definitions are connected with the purpose of the sentence, i.e. they include mention of the fact that "sentence" is a language unit serving to convey a "more or less complete thought". Only a few definitions are based on particularly formal criteria.

4. The defect of "sense" definitions lies chiefly in the fact that they violate the principle of homogeneity, they depart from the sphere of language as a system and assume or sanction the dissolving of an object of linguistics in an object of logic or psychology. Moreover, phrases like "more or less complete thought" and even simply "complete thought" are not defined more or less strictly in logic itself. The linguists are thereby doomed to waiting passively for the progress of logical semantics which it is easy to demonstrate, cannot itself develop without greater precision of linguistic concepts.

5. The defect of existing "formal" definitions as compared with "sense" definitions is that they lack the idea of syntactic coherence (according to Aidukevich), i.e. what is most important in this unit of language for a linguist.

Sentence "coherence" is reflected, as a rule, in the "sense" definitions, but it is reflected functionally through the coherence of the judgment.

6. The formal definitions of "sentence" /PREDLOZHENIYA/ coincide in substance with the definitions of "sentence" /FRAZY/. Meanwhile, the linguist is acutely aware of the need to distinguish between the two concepts.

7. The theory-of-numbers conception of language created by Soviet mathematicians is a completely explicit metalanguage of linguistics in which the basic linguistic categories may be rigorously defined.

8. In particular FRAZA /sentence/, i.e. the ordered succession of smaller units is taken as the original, undefined concept (the aggregate of meaningful or correctly constructed sentences in a certain language is considered given).

9. Introduction of the concept of configuration, strictly defined in metalinguistic terms, makes it possible to describe a relation of syntactic dependency, while the isolation of regular configurations enables us to obtain the complete analogue of a "syntagma" or "word combination".

10. The individual elements (parts) of the syntagma (they are described in the formal system as S-groups or relatemes /RELYATEMY/) may be regulated by the relationship of syntactic subordination. It is in these terms that the concept of coherence is formulated.

11. A sentence may be called a cohesive number of S-groups (or relatemes) such that for each S-group A there is one and only one S-group B such that B is syntactically subordinate to A.

12. Calling two S-groups linked by mutual syntactic subordination a predicative pair leads to the following theorem: a sentence has one and only one predicative pair.

13. The suggested definition meets all the requirements set forth at the beginning of this report: it is formal, reflects the idea of coherence, and is sufficiently close to the intuitive conception of the term "sentence". It also permits us to derive deductively the idea of predicativity.

14. Exclusion of the so-called "single-constituent sentences" is justifiable on two main grounds: first, whatever may be our definition of sentence, "single-constituent sentences" cannot in general be taken into consideration because the method of configurations is not applicable to them. Second, the problem of correlation of "single-constituent sentences" with a judgment cannot be completely solved. And it is important for us that the "sentence", determined by particularly formal means, may be placed in mutually well-defined congruence with the judgment. Thus, a strictly formal definition of the sentence is important even for logic.

26. TRANSLATION sub specie structuralismi

A. A. Reformatskii (Moscow)

1. Translation results from the variety of languages and the consequent lack of mutual understanding between their speakers. The purpose of translation is to supply necessary information (business, scientific, artistic, etc.) in language comprehensible to a given user of the information.

2. What is the "theory of translation" and can there be a special science of translation?

Criticism of "literary expansion" (L. N. Sobolev and, in part, I. Etkind). Where A. V. Fedorov is wrong in including the "theory of translation" in linguistics. The "theory of translation" not as a science, but as an object of science, even various sciences. The role of linguistics in the "theory of translation".

3. Types of translation. Narrowing of "scope of translation" in the usual view. Where L. N. Sobolev is wrong in considering translation limited to three types. How "type of translation is defined". A given text and the goal of translation. What is the structure of a given text in its known linguistic features and in its social trends. The linguistic features of a given text as determining the type of translation. Relevancy and non-equivalence of translation elements in various types of translation.

- 33 -

4. Translation as information and interpretation. What the structure of "translation" as a whole consists in. Initial data of translation, act of translation, results of translation in the structural sense. Where Z. Klemensevich and I. Etkind are right and I. Kashkin is wrong. Various types of relations between original and translation.

5. The problem of "translatability" and "non-translatability". What "lack of mutual understanding" consists in. Why Humboldt is right and Kashkin is wrong. The unwarranted claims of A. V. Fedorov and others. What is adequacy of translation in connection with analysis of translation elements and understanding of translation type.

6. Methods and circumstances of translation dictating various solutions of the translation problem. Ad hoc translations, translations are the "task of a lifetime"; lexicography, informative translations, technical and scientific translations, artistic translations, translations of philosophical texts; machine translation. Cooperation of sciences and talents in diversity of translation activity.

27. A SYSTEM OF RECORDING SPEECH FOR ORAL TRANSLATION

V. Yu. Rozentsveig (Moscow)

1. Oral translation differs from translation of a written document in that the words to be translated are perceived by the ear, transformed, stored in the memory, and later delivered orally. These operations take place more or less simultaneously (depending on the kind of oral translation).

2. The limited capacity of the "short" memory of man results in considerable losses of information when large segments of speech are translated. Moreover, overloading the memory makes the analysis and synthesis of a spoken communication difficult. It is necessary to work out a system of recording speech constructed in such a way that it would interact with the "short human memory, thus ensuring reliable storage of information, facilitating perception, and recreating the oral message.

3. A phonetic (alphabetic) writing system has not been devised for the recording of foreign speech. Stenography is unacceptable for oral translation because it registers the words in toto (including redundant and unnecessary words) and requires too much time to decipher. The system of shorthand worked out empirically in the University of Geneva's School for Translators largely meets the needs involved in recording speech for oral translation (Rozan's work). However, it does not solve our problem owing to its unsystematic nature and internal contradictions.

4. The task of developing an efficient system of recording speech for oral translation amounts to the creation of a unique elementary information language requiring the solution of several logical, psychological, and linguistic problems, to wit:

(a) logical analysis of speech, isolation of semantic fulcral points and systems of connecting them;

(b) identification of the properties and action mechanism of the "short" human memory;

(c) determination of linguistic redundancies, common (stereotypic) word combinations and sentences capable of being reduced to symbols, the most efficient techniques of designating morphemes and syntactic connections in the system of the complex whole. In addition, we must keep in mind the necessity of working out a recording system that will be applicable to a pair of languages and easily mastered by those studying to become translators.

28. LANGUAGE TRAINING FOR BLIND DEAF-MUTES

I. A. Sokolyanskii (Moscow)

1. The simultaneous lack of visual and aural analyzers and thereby of the speech analyzer is an exceedingly unusual condition for a child. The unusualness consists in the fact that the deaf-dumb-blind child is completely normal as far as neural and cerebral structure is concerned and therefore retains potentially the full capacity for intellectual development like that of any normal child. Nevertheless, using just his own efforts and without outside help he can not make initial contact with the external environment surrounding him.

2. Development of a deaf-dumb-blind child's first contacts with his environment is an extremely complex problem that can only be solved by selecting a rigorous system of initial signals. This is achieved by special teaching and a special grammar. Ordinary general (particularly "school") grammar, as presented in general courses cannot be used.

3. If the system of initial signal contacts is developed in close conformity with the logic of the external physical environment, formation of the second signalling system on the basis of the first is not particularly complex and is chiefly a technical problem. The heart of the matter lies, therefore, not in the second, but on the first signalling system.

4. The second signalling system (language) in teaching a deaf-dumb-blind child has various forms: gesticulatory, dactylic, touch (Braille), written, oral. The second signals must be strictly used in the same order as listed above.

A text is a basic link in the second signalling system--but no separate words or separate sentences. Hence, the language instruction of a deaf-dumb-blind child must begin with texts, not separate words or sentences.

-- 35 --

6. The beginning texts are short, consisting altogether of 3-4 so-called "simple" non-extended sentences (two-member). Five or six of these are enough, after which the child can pass to texts composed of "simple" extended sentences which, according to the rules, must include objects (direct or indirect indiscriminately).

The remaining syntactic constructions - even such difficult ones as complex clauses - are assimilated with the "simple" extended sentences in the series of texts. What general grammar calls a "simple" sentence is not at all "simple" as far as the teaching of deaf-dumb-blind children is concerned.

29. SOME GENERAL PRINCIPLES IN COMPILING GLOSSARIES FOR MACHINE TRANSLATION

G. M. Strelkowskii (Moscow)

1. The word -- as a basic unit of language. "Every word (speech) generalizes" (Lenin Philosophical Notebooks). Since ideas originate simultaneously with words and are expressed through words. The very possibility of logical thought is created solely by language. The unity of language and thought is organic, i.e. language can neither arise nor exist without thought, nor thought without language. However, words are not identical to ideas. Words may have several meanings, i.e. they may express different ideas and, vice versa, one idea may be expressed by several words. A word may contain not only the expression of an idea, but also the relation of the speaker to the object designated by the given word (KHOLOD /cold/, KHOLODISHCHE /extreme cold/ KHOLODOK /slight cold/etc.)

2. In this connection one should mention the impossibility of describing language without referring to meaning (the weakness in the theories of American structuralists and their followers). The unsoundness of theories reducing language to a system of pure relationships (Yel'mslev).

3. In accordance with the considerations set forth above, algorithms for machine translation must be based on a dictionary of meanings.

4. The principles of word choice for a machine dictionary.

    (a) Significant and auxiliary words.

    (b) Division of significant words into technical terms and words in common use.

    (c) Need to ascertain the minimum of international words required for comprehension of technical texts.

- 36 -

(d) Choice of subjects (Electronics, particularly the section dealing with automatic control, since this field is now included in all branches of industry and science, and is a basis for machine translation itself. Competence of author).

(e) Problem of compound words and word formation. Regular and irregular translation of compound words. Glossary of stems and program therefor or information referring to tables of suffixes and paradigms of word changes (including stem changes, e.g. stem forms of strong verbs).

5. Word combinations and phrases. Providing words in the glossary with an index indicating possible stock phrases. Translation of lexical homonyms by the method of analyzing word combinations.

6. Methods of work in compiling dictionaries. Choice of articles, reading them, writing out all words, except the commonest helping words (auxiliary verbs, pronouns, prepositions, etc.) oN index cards; alphabetic arrangement of cards. Numbering sentences in the text and corresponding index on the cards for ready location of possible occurrences of the word.

7. Statistical conclusions. Alphabetic arrangement of words. Percentage of technical terms. Repetitiousness of non-technical terms.

8. Methods of expanding the glossary with and without the machine.

Reading of other materials on the given subject and enrichment of glossary with common words.

Inclusion within the glossary of all technical terms already selected in the special glossaries of technical terms on the given subject. Treatment of new texts by the machine with separation of words not known to it and presenting them untranslated, or simply a selection of new words.

9. A selected glossary as a foundation for constructing a translation algorithm without the creation of some metalanguage.

30. SOME ANALOGIES TO THE PROBLEMS AND METHODS OF CONTEMPORARY THEORETICAL LINGUISTICS IN ANCIENT INDIAN GRAMMATICAL WORKS

V. N. Toporov (Moscow)

1. Linguistics has perhaps never been so independent and complacent as it is today. This is undoubtedly due to the fact that the real object of the science has been found. On the other hand, the connection between linguistics and other sciences has never before been so strongly felt.

But this connection is effected not on the earlier basis, when attempts were made to apply the methods of one science to another, but on a new one. It is characterized by some ideas common to a number of sciences. These ideas developed (often independently) on the soil of the various sciences. The isomorphism of certain fundamental concepts (cf."structure", "field", "invariant", etc.), the similarity of individual problems and methods of solution. It is becoming increasingly evident that certain common ideas and methods are being superimposed, as it were, on the material of the particular sciences and transformed in accordance with the nature of the material, the possibility of giving it a strictly formal inter- pretation, the scientific traditions in the given field, etc. For this reason the prospects for a new synthesis of various sciences on a new basis are now being carefully assessed (cf. International Encyclopedia of Unified Science, vol. I, 1938-1939; B. Hanssen. The concept of field as a synthesis of natural science and humanities traditions in sociology. Vestnik istorii miroroi kul'tury /Herald of the History of World Culture/, 1957, no 4, etc.).

At this time when linguistics is very clearly aware of its place among the other sciences and the new direction in linguistics is inter- preted as being something broader than simple opposition to old ideas, it is natural that there should be growing interest in the outlook for the development of linguistics, the nature of its connections with other sciences, and the ultimate fate of these connections.

When one examines these problems, it is difficult to avoid thinking about certain striking analogies to modern linguistic problems that may be found in the history of ancient Indian science, particularly linguistics, and which are attracting the attention of modern scholars with increasing frequency (L. Bloomfield, Emeneau, Bro, Allen, Renov, and others).

3. Let us list the most important analogies in the light of con- temporary problems.

(a) Formal principle of language description ("descriptivism"), exclusion of meaning in analysis, if we disregard the very small number of Sutra-interpretations that sometimes deal with the determination of connections of semantic (according to Morris) order; fullness of description, including differentiation between the obvious and the non- obvious).

(b) Elements of a systematic approach to language; clear destinction between class and member of class with fixed place; hence, on the one hand, the concept of zero, on the other, potential forms, hypergrammaticisms, false variation (often supported by the striving for conciseness in exposition); contrast of Sphota-sabda; negative character- istics of members in relationship; Prabhakara's teaching on semantics-- schools on relation of word and sentence and the dependence of the former on the latter; distinction between signum-designatum-denotatum.

(c) The metalanguage of Indian grammatical treatises; symbols (sign-index, sign-symbol types); metalanguage grammar in cognition.

(d) In connection with these features of ancient Indian linguistics, mention must also be made of similar phenomena in other fields: The esthetic code in ancient Indian art, particularly in the drama; the concept of dhavani (an analogy to sphota); some analogies in the works of ancient Indian logicians and philosophers (categories of relation, time; "nominalism"); characteristics of Indian historiography; the concept of zero among the mathematicians of ancient India, etc. A comparisn with ancient Greek science enhances the significance of the specific features of ancient Indian grammatical literature, which in many respects resembles modern linguistics.

31. THE FREQUENCY OF LEXICAL UNITS IN ENGLISH GEOLOGICAL LITERATURE

M. G. Udartseva (Petrozavodsk)

1. We undertook a study of frequency of lexical units in English geological literature in connection with the compilation of a minimal glossary for students in geological institutions. As material we selected articles on the various branches of geology as well as on the allied sciences. In addition, for the sake of objectivity in the tally, we included a considerable number of authors from several English-speaking countries. The final listing of sources comprised 33 works containing a total of 250,000 words, of which 28 are articles from 14 periodicals published in the United States, Great Britain, Canada, India, and Australia, while 5 were excerpts from monographs.

2. The literature dealt with problems related to the following branches of geology: mineralogy, crystallography, petrography, petrography of sedimentary, igneous, and metamorphic rocks, petrology, stratigraphy, paleontology, lithology, tectonics and structural geology, origin, distribution, and exploitation of mineral resources, geology of oil and coal deposits, geophysical methods, prospecting for mineral deposits, radioactive methods of determining the age of rocks, quaternary geology, geomorphology and glaciology, dynamic geology, geology of the ocean bottom, and regional geology.

3. Individual words, phrases, and verbs plus post positions were used in the count. Each additional meaning of a word was handled as a separate item. For example, the word "face" was regarded as four separate words corresponding to the meanings of "side", "face" (of crystal), "surface", "to put something in front of a person".

4. Each lexical item encountered again was entered on a separate index card where all secondary usages with indication of author were noted. If the word occurred more than 100 times in different authors, no further entries were made. Such words as "that", "which", "it", etc. were handled similarly.

5. The count resulted in a determination of the frequency for 7535 words. We entered into the minimal glossary 2373 lexical items consisting of 546 verbs, 954 nouns, 327 adjectives, 236 adverbs, and 310 other kinds of words. Of this number 176 words are specialized terms; more than 200 words have another meaning in geological literature, while the remainder are ordinary words. About 4000 of the 7535 words are technical terms.

6. The minimal glossary was tested by taking several random pages of diverse literary, general political, and geology material and cal-culating the percentage of words from each text that were lacking in the minimal glossary. It turned out that a page of geological text contained 1-1.5% "unfamiliar" words, general political text 8-10%, and literature (Dickens) 16-18%.

7. The minimal glossary was also collated with the Thorndyke dictionary. Significant discrepancies were noted even in determining the first 500 words.

32. ONE APPROACH TO LOGICAL SEMANTICS

V. K. Finn and D. Kh. Lakhuti (Moscow)

1. Our approach to logical semantics can be summed up as follows:

   (a) some language of science with minimal pragmatics is selected as the investigated language (e.g., the language of synthetic organic chemistry, formal genetics, classical mechanics, etc.);

   (b) an artificial language is constructed for the investigated language I and it consists of a glossary (class of basic technical terms and syntactic functors) and a class of indexes for the glossary as well as a formal syntax in which are for-mulated the rules for building sentences consisting of the indexes. A correctly formed sentence in language I is determined with the help of an algorithm constructed in the formal syntax.

   (c) Language I is expanded into language II consisting of language I. A list of descriptions of types of sentences in language I (examples of such types of sentences for the language of synthetic organic chemistry will be: sentences conveying in-formation about compounds; sentences conveying information about reactions; sentences conveying information about re-action conditions) and a list of combinations of indexes cor-responding to the types of sentences formed.

(d) In accordance with the types of sentences algorithms are constructed in language II that discern the meaning of these sentences.

   If sentence F is correctly constructed and all the indexes are replaced by dictionary signs and if the combination of indexes corresponding to F coincide with the combination of indexes of some of the sentence types in language I, the algorithm will convert F into sign "S", if all the predicates of the corresponding description are satisfied for F; if even one predicate of the description is not satisfied for F, the algorithm will convert F into an empty word.

   In the first case we will say that "F has meaning in language I", in the second "F does not have meaning in language I". If, however, the algorithm is not applied to F, we will say that the meaning of F is not determined in language I.

   A descriptive syntax is formulated in language II. It consists of suitable algorithms to discern the meanings of sentences and a list of rules according to which meaningful sentences are derived from meaningful sentences.

(e) Language II is subsequently expanded into language III in which definitions with reference to the properties of language I and its relations to the investigated language are formulated. Language III consists of language II and a list of definitions.

   Language III contains definitions of the concepts of the semantic completeness of language I, translatability (full or partial) of the investigated language into language I, interpretation of language I within the amalgamation of language II and the investigated language, explicitness of language I, and other semantic concepts.

   If it is possible to construct a series of languages I, II, III for the investigated language, we will say that the "semantic analysis of the investigated language" has been realized. If the investigated language is at least partially translatable into language I, it is suggested that "semantic analysis of the investigated language" can be effected by an automatic machine.

(f) "Semantic analysis" is in the experimental stage, and that is why we speak about an "approach" to semantics, and not the construction of a deductive system of semantics.

   However, the deductive construction of a system of semantics is possible on the basis of experimental investigations of the "languages of science" (with minimal pragmatics).

In the preparation of this paper we have used the ideas and results of research in semantics by A. Tarski, K. Aidukevich, L. Hwistek, I. Bar-Hillel, G. Curry, V. Quine, N. Chudman, and R. Carnap.

33. SOME PROBLEMS CONNECTED WITH THE HANDLING OF VERBS WITH ALTERNATING STEMS IN CONSTRUCTING AN ALGORITHM OF MACHINE TRANSLATION FOR SPANISH

(A Statistical Inquiry)

R. M. Frumkina (Moscow)

The compilation of a dictionary of stems is a necessary stage in the task of constructing an algorithm of machine translation. By stem we understand the graphically invariant part of a word. However, there are a number of languages in which the graphically invariant part of certain words, principally verbs with alternate forms, consists of one or two letters, an inconvenience resulting in homonomy of stems. It is therefore necessary to separate only the purely standard endings (person, number, etc.), and assume that a given word has several stems.

There are two possible ways of solving the problem:

(1) Enter into the dictionary all the stem variants of each word with plural stems, e.g. perfective and imperfective aspect, present and past tense stems, etc. We thereby increase (and sometimes considerably) the size of the dictionary.

(2) Select the most frequently occurring variants and enter them into the dictionary; for the other stems, furnish the rules by which they are in some manner to be identified or formed according to the stems listed in the dictionary. This would enable us markedly to reduce the size of the dictionary, but at the price of complicating the program.

In order to determine the more efficient method, it will be necessary above all to carry out a statistical inquiry concerning words with plural stem variants and their frequency. We are now analyzing the frequency of verbs with alternating forms in a Spanish scientific (mathematical) text. On the basis of data in the frequency dictionary of V. Garcia Hoz, all Spanish words with a frequency of more than 40 were first divided into classes depending on the types of alternation. Then the frequency both of classes and of individual morphological forms was determined from consecutive material in mathematical texts.

The data thus obtained clarify the principles governing the distribution of classes and alternating forms and enable us to make certain recommendations in compiling a dictionary and rules for handling stems.

34. A LOGICAL ANALYSIS OF THE CONCEPT OF LANGUAGE STRUCTURE

S. K. Shumyan (Moscow)

1. Modern structural linguistics interprets language structure on the Gestalt plane, ie, as a whole, the elements of which are connected by definite relations.

2. If we consider that language elements interact on two axes-- syntagmatic and paradigmatic--an interpretation of language structure on the Gestalt plane must be regarded as one-sided: we encounter wholes, the elements of which are connected by definite relations, only on the syntagmatic axis (such wholes, for example are syllables in phonology or sentences in grammar. However, on the paradigmatic plane we deal not with wholes, but with classes of ordered elements: the elements of these classes are interlinked by definite relations, but the classes can not be identified in any way with the wholes.

3. There arises the need of defining language structure in such a way that the definitions may be applied to the interaction of language elements not only on the syntagmatic, but also on the paradigmatic axis.

4. The new definition of the concept of language structure is based on the general concept of structure in modern symbolic logic where it is defined thus: the structure of a given relation is the property of being isomorphic with the given relation.

Modern structural linguistics, as we know, distinguishes two planes in language: the plane of expression and the plane of content (phonology is included in the former, grammar and lexicology in the latter). Since isomorphism exists between both planes, we may rely on the definition of the general concept of structure in symbolic logic and define language structure thus: language structure is the property of the relations of elements on the plane of expression and of the relations of elements on the plane of content to be isomorphic with one another. This definition of language structure is in complete accord with the research techniques of structural linguistics at its present stage of development.

6. A logical analysis of the concept of language structure requires an operational approach to this concept. Accordingly, the report states how we should set up empiric operations by means of which language structure, as an abstraction, can be linked to genuine linguistic activity.

35. ANCIENT TEXTS AND MACHINE TRANSLATION
(A formulation of the problem)

V. Shevoroshkin (Moscow)

1. There is no doubt that a great many philosophers, historians, ethnographers, and even specialists in literature have an acute need of Russian translations of a large number of ancient texts.

2. The available translations are a drop in the ocean compared with the mass of ancient literary monuments.

3. Texts in dead languages have one feature that distinguishes them from texts in modern languages, namely, the frequent impossibility of proving that the original author had in mind precisely what we "read into" the text.

4. The feature of ancient texts noted above has produced and is continuing to produce numerous commentaries on these texts.

5. The translator of ancient texts is in essence a commentator. Even the translator who strives for maximum objectivity inevitably introduces into his work many subjective elements, which vary in degree with the depth of his erudition.

6. An investigator who requires the translation of an ancient text may also need a commentary, but his primary need is for a maximally objective translation. When reading such a translation, he should confront the same difficulties that are mastered by a person who reads the text in the original. However, a translation done by a human being does not meet these needs for the reasons mentioned in (5) above.

7. Machine translation of ancient texts will enable a student to obtain exactly what he needs. "Interpretation" of a text by a machine is excluded. The more "elementary", the better.

8. Thus, machine translation of ancient texts is particularly important, for the machine is not merely a substitute for a live translator, but - in this respect alone - it does what a person can't do.

9. Certain characteristics of the ancient Indo-European languages enable us to assert that these languages are more accessible to machine translation than are the living languages. These characteristics include: Comparatively greater transparency of morphology and simplicity of syntax, numerous trite phrases, etc. This problem will be considered in detail on Sanskrit material.

10. For the reasons set forth above machine translation of ancient texts into Russian is a problem that deserves detailed elaboration.

SECTION ON ALGORITHMS OF MACHINE TRANSLATION

36. AN ALGORITHM FOR TRANSLATING FRENCH INTO RUSSIAN
ELECTRONICALLY

V. A. Agrayev
(Gorki)

The algorithm was designed for use in connection with an electronic computer of the GIFTI /Gor'kovskii issledo-vatel'skii fiziko-tekhnicheskoi institut/Gorky Research Institute of Physics and Technology/possessing a limited memory capacity. The aim was to determine the translation capabilities of the machine as well as to check the operation of the algorithm with limited glossary and rules.

The algorithm includes lexical routines: a glossary of stems, a glossary of phrases, and charts for translating polysemants. The stem glossary contains about 500 words. In addition, we prepared a large glossary (about 1200 words) containing the full, original forms. The amount of grammatical information included with the words varies in the two glossaries: less is given in the stem glossary. The phrase search is based on the semantically pivotal word. The translation routines of polysemants contain tests for contextual environment and the required meaning is selected accordingly.

Analyzing rules determine the meaning of French inflections and, depending on the governing words, establish the necessary grammatical forms of the other words.

In the synthesis routines Russian word forms are constructed on the basis of grammatical information derived from the glossary and developed during the process of analysis. Synthesis is effected with regard for its applicability also to translating English radio engineering texts.

Statistically chosen data were used in constructing the algorithm.

37. PRINCIPLES IN THE CONSTRUCTION OF ELECTRIC
READING DEVICES

N. D. Andreyev
(Leningrad)

1. The problem of electric reading devices (EChU) /elektrochitayushchiye ustroistva/ arises because of the slowness in preparing a text for machine translation, which is inevitable when a human being does this work (particularly in oriental language texts).

2. An electric reading device must be adapted for machine sensing of scripts of varying size, slant, proportion, and graphic shape.

3. The different sizes, slants, and proportions of scripts may be reduced to a single standard by using the three-set system of varying curve

- 45 -

mirrors /TREKHKOMPLEKTNOI SISTEMY ZERKAL PEREMENNOI KRIVIZNY/.

4. Scripts of different shapes may be adapted for machine sensing
by using the principle of key identification points /KLYUCHEVYKH
OPOZHAVATEL'NIKH TOCHEK/, the number of which cannot exceed 50 for Cyrillic
and Latin; it may reach 100 for Arabic, Devanagari and their derivatives,
and about 300 for Chinese and Japanese.

5. The set of key points is individualized for each of the graphic
signs and is interpreted for each language in accordance with a special
program that constitutes the introductory part of the analysis in the
appropriate algorithm.

38. WORK ON AN INDONESIAN-RUSSIAN ALGORITHM
OF MACHINE TRANSLATION

N. D. Andreyev
(Leningrad)

1. The Indonesian language requires preliminary treatment of the
words in order to strip their roots. Stripping of the root by direct
resort to a dictionary appears to be impossible.

2. Three factors make it difficult to strip the root; (1) the
presence of initial and secondary prefix and suffix; (2) internal sandhi,
i.e. the phonetic interaction of morphological elements; (3) the presence
of root reduplicators and polyreduplicators, which occur in two graphic
variants.

3. Much preliminary work was required for the statistical and
structural investigation of Indonesian words. Different versions of the
root-stripping program were based on this work.

4. Processing the words in the root-stripping program makes it
possible to proceed to morphological analysis, which is effected by
a special morphological program that is often realized in a purely
analytic way, i.e., without resorting to the output language, but by
substituting words in their code hieroglyphic.

5. Based on a certain working hypothesis concerning the structure of
the Indonesian sentence, it seems possible to construct a standard analysis
constituting the principal part of the syntactic program; it is only for a
minor portion of the sentences that we need a non-standard analysis forming
a more complicated but much less frequently used part of this program.

6. The homonym and phraseology programs are operated after the first
three programs are completed, relying on the hieroglyphic analysis effected
therein.

7. The propositional and glossary program works chiefly by conversion, i.e., according to the output language.

8. Tables of pseudoroots and typical sets of morphological information are being developed as necessary supplements to the main glossary.

39. WORK ON A VIETNAMESE-RUSSIAN ALGORITHM OF MACHINE TRANSLATION

N. D. Andreyev, D. A. Batova, and
V. S. Panfilov (Leningrad)

1. The Vietnamese-Russian algorithm of machine translation includes the following programs:

(a) Glossary of binomials, /BINOM/

(b) Glossary of roots,

(c) Glossary of idioms,

(d) Supporting program, /OPORNAYA PROGRAMMA/

(e) Syntactic program,

(f) Homonymic program.

2. The glossary of binomials assumes the stripping of two-syllable Vietnamese words with their grammatical information.

The glossary of roots includes monosyllabic words and their grammatical information. The existence of two glossaries is due to the problem of word boundary in isolating languages.

The glossary of idioms contains idioms, phrase combinations, and hard-to-translate expressions.

The supporting program serves to differentiate between parts of speech in those cases where the appropriate grammatical information cannot be precisely indicated either in the glossary of roots or in the glossary of binomials.

The syntactic program provides for an analysis of Vietnamese syntactic constructions.

The homonymic program is designed to solve the problem of lexical homonymy within any single part of speech. The program deals principally with monosyllabic words, since homonomy is not characteristic of binomials.

3. In connection with the adoption of a syntactic standard, which consists in utilizing syntactic analysis to determine the parts of speech, the range of application of the supporting program is narrowed to exceptions to standard cases.

4. Besides utilization of the supporting program, exceptions to standard cases may be solved by inserting appropriate corrections into the syntactic program.

5. The supporting program is characterized by:
        (a) The ability of individual words to occur in a sentence as a substantive and a verb.

        (b) The fact that such words stand closer to the verb than to the substantive. Therefore, when used as substantives, they often receive various grammatical indicators that are peculiar to substantives.

        (c) A number of verbs may be brought into the category of substantives by means of appropriate auxiliary elements.

        (d) What has been set forth above explains the impossibility of accurately indicating in the glossary the part of speech of the words in question. The part of speech may be indicated only disjunctively.

        (e) Determination of the part of speech to which words of the type in question belong may be made in each specific case with the help of carriers of grammatical data located in the supporting program.

40. WORK ON A JAPANESE-RUSSIAN ALGORITHM OF MACHINE TRANSLATION

A. A. Babintsev
(Leningrad)

1. Work on a Japanese-Russian algorithm was begun at the end of December 1957, using atomic energy texts. At this stage analysis of material is limited to the simple sentence.

2. Due to the fact that no reading devices are available for ideographic text, the Japanese must be transcribed into Russian before it is put into the machine.

3. The structure of Japanese--agglutination (substantive and verb in part) and inflection (verb in part and adjective) with the stress on agglutination--is responsible for the effectiveness and adequacy of the standard morphological analysis and determines the primacy of the program of standard morphological analysis in the set of programs.

4. The set of programs for the Japanese-Russian algorithm at the present time is as follows:

(1) A program of standard morphological analysis (with referral to the glossary--"address" and withdrawal therefrom of certain grammatical information).

(2) A program of standard syntactic analysis (based on a "working hypothesis").

(3) A program of non-standard syntactic analysis (cases that do not fit the "working hypothesis").

(4) A homonymic program.

(5) A glossary of idioms.

(6) A synthesizing program.

5. The minimum of information to be derived from text analysis is: for a substantive--case and, in certain instances, number; for a verb--tense, voice, mood, finiteness; for an adjective--tense.

6. The "working hypothesis", which is based on the laws of Japanese sentence structure, in broad outline consists of the following:

(1) The first substantive in the nominative or principal case is the subject.

(2) The last word before a stop sign is the final predicate; a verb in non-finite form is the middle predicate.

(3) The direct object immediately precedes the predicate; the indirect object is found at some distance from the predicate.

(4) A substantive in the genitive case, adjective and verb in the finite form preceding the substantive are attributes.

7. We should like to direct attention to one of the numerous problems that have arisen in connection with our work on the algorithm. After analyzing a Japanese text, from which information on number can be obtained only sporadically, it turns out that difficulties due to the inadequacy of information on grammatical number appear in the synthesizing program during formation of the output text. A solution to the problem of number in the synthesizing program is exceptionally important for a number of "oriental"-- Russian algorithms.

41. THE PROGRAMMING OF TRANSLATION FROM ENGLISH INTO RUSSIAN

G. P. Bagrinovskaya and
G. L. Gavrilova (Moscow)

Program of translation, constituent parts, order of operation.

Arrangement of glossary, difference in coding used in English section of glossary from coding in French section of glossary. Size of glossary. Glossary of phrases.

Choice of homonyms, construction of complex index scales and omitted index scales.

Operation of analysis program ("rolling up" formulas) /FORMULY SVERTKI/. Program of synthesis of structures on the basis of formulas of synthesis. Morphological treatment of results of synthesis.

Russian part of program of translation from English into Russian (utilization of programs prepared for Russian part of French-Russian translation). Agreement in codings.

42. PRINCIPLES IN COMPILING A GERMAN-RUSSIAN GLOSSARY OF POLYSEMANTS FOR MACHINE TRANSLATION

S. S. Belokrinitskaya (Moscow)

Determination of the meaning of a polysemantic word that is appropriate in a given context constitutes one of the basic problems in machine translation.

This problem is being solved by compiling a glossary of polysemants which will make it possible to obtain the relative meaning of a word by an analysis of the surrounding context. In most cases it is sufficient to examine context within the boundaries of a sentence.

A considerable number of words that have multiple meanings in the usual literary language have but a single meaning in mathematical texts, and the system of meanings for a number of polysemants is simplified. However, many German words, even in a mathematics text, have a large number of relative meanings, the determination of which requires a rather complicated system of tests.

The most numerous are prepositions and a group of verbs which are used with separable prefixes and which also form a large number of phrases.

The principal method of determining the relative meaning of a polysemant is structural--semantic analysis of the surrounding context. In some cases grammatical forms of the given word or its environment are also analyzed .

It is possible to isolate certain groups with a monotypic system of meaning, thereby simplifying the glossary and replacing in some cases the system of tests (or part of the system) by reference to the appropriate general rule.

We have also isolated a group of words united according to the principle of identical effect on the translation of prepositions and some verbs with extremely many meanings, which likewise permits of simplification of the routine.

Methods of glossary treatment of different types of idioms and phrases have been worked out.

The routines of polysemants also contain cases of lexical homonomy that are not excluded from the system that differentiates between the meanings of polysemants.

The determination of relative meanings of polysemants by means of the glossary just described is not free from difficulties (in some cases a single sentence does not provide sufficient context, the translation of complex words, etc.). However, these difficulties can, as a rule, be overcome.

A check of the text shows that a complete satisfactory translation of the mathematical corpus can be achieved with the help of the above-described glossary of polysemants.

43. MAIN FEATURES OF THE GLOSSARY AND GRAMMATICAL PROGRAMS FOR ENGLISH-RUSSIAN MACHINE TRANSLATION (M.T.)

I. K. Bel'skaya (Moscow)

1. The basic components of a system of machine translation from English to Russian as worked out in the ITM* and VT*, /*See No. 2 for expansion and meaning of abbreviations/Academy of Sciences, USSR are a specialized bilingual glossary and three cycles of translation routines: glossary routines, routine for analysis of imput sentence and routine for synthesis of output sentence.

2. The Anglo-Russian M.T. glossary now available has been designed for the translation of scientific literature dealing with problems of

applied mathematics: the solution of systems of linear, algebraic, and transcendental equations, calculation of the proper values of matrices, approximation of functions by means of polynomials as well as by trigonometric functions, expansion of functions into series, numerical differentiation and integration, numerical solution of differential equations, and other problems of numerical analysis.

The glossary contains 2300 words. Several works by English authors were used for compilation and checking.

Text checking of the glossary for translation of mathematical literature yielded satisfactory results. Some 3000 sentences consisting of more than 100 connected passages from the material of different authors were used as the corpus.

3. A glossary for the machine translation of scientific literature may be usefully divided into a series of independent "specialized" glossaries. Further specialization down to relatively independent fields within a given science--mathematics, physics, and chemistry--is also worthwhile.

This division serves two purposes: it reduces the necessary bulk of the glossary to the completely manageable number of 3000-3500 words and even more important, considerably reduces the amount of polysemy.

The structure of the Anglo-Russian glossary for M.T. is such that its several sections may be expanded independently.

The glossary has two main sections:

  I Single-meaning glossary and

  II Multiple-meaning glossary.

Each section is divided into two subsections:

  Ia - glossary of terms,

  Ib - glossary of words in general use,

  IIa - glossary of words with complete meaning,

  IIb - glossary of auxiliary words.

In size, the multiple-meaning glossary takes up about 1/5 of the entire glossary which, in this instance, amounts to 458 words.

5. The problem of polysemy is satisfactorily solved by combining two methods: (a) narrow specialization of a series of glossaries for M.T. and (b) contextual (functional-semantic) analysis of words in the sentence. Experience shows that it is virtually unnecessary in scientific and technological texts to go beyond the "small context"(i.e. one sentence).

6. In order that the lexical analysis of the words be effected automatically (without human intervention), the M.T. glossary is accompanied by a series of special glossary routines that make up cycle I in the over-all system of translation routines.

These include:

1. A routine for obtaining the glossary form of the words,

2. A routine for the grammatical analysis of "unknown words",

3. A routine for the grammatical analysis of "formulas",

4. A routine for distinguishing homonyms,

5. A routine for the analysis of polysemy.

The last routine is the most important from both the theoretical and the practical points of view.

7. The lexical analysis, which is performed by means of the glossary and glossary routines, precedes the grammatical analysis and provides it with the necessary initial information in the form of the so-called "invariant characteristics" of each "known" word (i.e. entered in the M.T. glossary) and the syntactic characteristics of all the "unknown" words (not entered in the M.T. glossary) and the "formulas".

8. The grammatical analysis of input sentences is performed by means of a series of routines in cycle II in the following order:

1. Analysis of verbs ("verb" routine);

2. Analysis of punctuation marks;

3. Syntactic analysis of sentences: division of sentence into clauses and more precise definition of parenthetical phrases in clauses/We define a sentence as that segment of text which is included between full stops (period, exclamation or interrogation point); a clause is a simple sentence, i.e. such that it contains no more than one heterogeneous predicate./;

4. Analysis of substantives and numerals;

5. Analysis of adjectives;

6. Modification of word order in the translated sentence.

The "verb" routine is the key routine in the first half of the analysis of English sentences; however, the syntactic analysis of sentences (routine 3) is the basis of operation for the second half of the analysis and determines the boundaries of those segments within which the subsequent analysis is effected.

9. The routines in cycle III use the results of the preceding routines in such a way that the Russian sentence obtains its grammatical form in accordance with the rules of Russian grammar.

The synthesis routines go into operation just at the time when the variant (contextual) grammatical signs for all variable words in the output sentence are obtained and the steps taken to adjust the word order to Russian norms.

In the place of the Russian numbers, which represented Russian words up to this time, Russian equivalents are selected from the glossary, after which the variable words (verbs, substantives, numerals, and adjectives) are handled by the synthesis routines: a word ending is changed whenever the desired word form does not coincide with the dictionary form of the word.

10. Synthesis routines operate in the following order:

1. Word-forming routine;

2. "Verb" routine;

3. "Adjective" routine;

4. "Substantive" routine.

Changes in the numerals are effected partly in the "substantive" routine, partly in the "adjective" routine.

The word-forming routine occupies a special place: it provides for various cases going beyond word changes while inserting the grammatical signs of the Russian word derived from analysis of the foreign sentence.

44. WORK ON A NORWEGIAN-RUSSIAN ALGORITHM OF MACHINE TRANSLATION

V. P. Berkov (Leningrad)

I. The projected set of programs are: A. Analytic part: (1) morphological program; (2) program for distinguishing homonyms; (3) syntactic

program; B. Glossary part: (4) glossary--address; (5) regular glossary; (6) glossary of phrases and idioms; (7) program for compound words; (8) prepositional program; (9) program for unification of orthography; C. Synthesizing part.

II. Two methods of analysis, different in principle, were initially contemplated:

(a) To begin with a search for words in the glossary;

(b) To begin by extracting grammatical information from the text before referring to the glossary on the basis of a supporting program (lists of indisputable endings, word-forming suffixes, supporting words, etc.). Due to the extensive amount of grammatical homonomy in Norwegian, the second method seemed very cumbersome and, in some cases, practically unsound. It has therefore been rejected.

III. The fact that the functioning of the algorithm--which begins with a search for the words in the glossary and withdrawal of the information located there into the operative memory--leads to clogging the latter with information that is as a rule temporarily superfluous (in some cases this is general) suggested the idea of creating typical sets of information.

IV. Programs (1) and (2) are now (beginning of March 1958) ready in rough form. An ending obtained by stripping the dictionary stem from the text form of the word is compared with the list of endings; if the given word has a single grammatical meaning, an information suffix is attached to it and no further action is taken on the word at this stage. Cases of grammatical homonomy are handled by a series of special programs (2). On extracting all the grammatical information from the text linear transfers of words are made in order to impart a standard appearance to the items derived by "unrolling" /RAZVERTKI/: this is done by a part of program (3).

V. The program for the unification of orthography is the specifically Norwegian part of the algorithm. The need for this program is dictated by the considerable amount of inconsistency in Norwegian orthography, even in scientific texts; without the program, the glossary would necessarily be overloaded with many pairs of words.

VI. The program for unification of orthography will be used as a basis on which to construct an adjusting program in connection with the use of this algorithm for Danish.

45. GLOSSARY STRUCTURE AND INFORMATION CODING
FOR MACHINE TRANSLATION

I. L. Bratchikov, S. Ya. Fitialov,
and G. S. Tseitin (Leningrad)

1. Consideration is being given to the problem of introducing a
glossary on tape into the machine to search for coincidences in the event
that the glossary does not fit into the operational memory.

2. A glossary structure is proposed that will accelerate searching
and decrease the size of the indispensable portion of the memory for the
size of the glossary under consideration.

3. The previously suggested process of "rolling up the codes"
/SVERTYVANIYA KODOV/ is now in use. The rolled up code is directly
utilized to obtain the address of information on words in the glossary.
We have provided for cases of coincidences of addresses thus obtained
(rolling up SVERTOCHNAYA/ homonomy) by differentiating routines in-
cluded in dictionary compartments, the addresses of which are not
addresses of the words.

4. Theoretical probability considerations have enabled us to obtain
results which, based on the given number of words in the glossary and the
volume of lexical information, make it possible to estimate the necessary
size of the memory to accomodate the glossary.

5. Methods are also suggested for programming certain operators
encountered in the algorithms of machine translation.

46. GENDER AS A SUPERFLUOUS CATEGORY OF
THE RUSSIAN VERB

V. N. Vinogradova (Moscow)

It is very important for machine translation to discover the gram-
matical categories of a language concerning which there is no need to
give information insofar as translation can be effected without taking
them into account. Certain general considerations suggest that gender
in the Russian verb--an uncharacteristic phenomenon expressed only
in —1 forms, the singular of the past tense and of the conditional mood--
is one of these categories.

We tested this assumption on a mathematical /I. G. Petrovskii,
Discourses on the Theory of Differential Equations, 1954/text where the
number of verbs with gender expressed turned out to constitute only 4%
(93) of the total number of verbs (1970). We then selected linguistic
(history of language)/A. A. Shakhmatov, Historic Morphology of the
Russian Language, 1957, pages 9-65/ and historic /B. D. Grekov, Kievan
Russia, State Publishing House of Political Literature, 1953, pages 73-
142/ texts in order to have a large number of diversified examples and
found that these verbs constituted about 30% of the total number of verbs

used. It appears that in most sentences the verb may be related only to the subject--a single substantive in the nominative case. Doubts may arise only in the case of transitive verbs where there is an object in the accusative case that coincides in form with the nominative, of the type: "Equation (6) yielded the general integral of this equation over the entire surface except for the start of the coordinate". /Uravneniye(6) davalo obshchii integral etovo uravneniya vo vsei ploskosti 3a isklyucheniyem nachala Koordinat/. Since we have a grammatical indication for both the subject and the object purely in the past tense, and even then only when the gender of the noun-subject differs from that of the noun-object, there remains no other way of determining which is which than the word order of the sentence: the rule that the subject comes first holds in the over-whelming majority of cases. A rearrangement is, of course, possible for the sake of logical emphasis, e.g.: "in the Russian language preponderance has received the accent of the nominative plural." /V russkom ya3yke pereves poluchilo udareniye imenitel'novo mnozhestvennovo/. The phrase poluchit' pereves /receives the preponderance = predominates/ will evidently have to be listed in the glossary as a phrase combination.

It is possible to conceive of more complicated cases (we didn't find any examples, but we paraphrase one of the sentences of the type described above): "Change...caused a shift of e to o before a hard con-sonant." /Izmeneniye...vyzval perekhod e v o pered tverdoi soglasnoi/ Such a sentence is almost impossible with a predicate in the present tense (or is very badly written: "Izmeneniye ...vyzyayet perekhod..." will clearly be misunderstood); even in the past tense it is awkward. Apparently, rare instances of this kind will be edited; so too the following case in a complex sentence: "The bishop asserted that his church land went along the Lisichii ford, which was in the time of Prince Yuri." /Episkop utverzhdal, chto yeo tserkovnaya zemlya idet po Lisichii brod, chto byl pri knyaze Yurii/. In the absence of information on the gender of the verb byt', it is impos-sible to determine whether the last clause modifies brod "ford" (brod, chto byl...= kotoryi byl... an obsolete meaning, according to Ushakov's Tolkovyi Slovar' /dictionary/) or is a subordinate conjunctive clause relating to the entire preceding clause. This ambiguity cannot be resolved here by formal signs.

With the exception of the last example, the texts studied did not con-tain a single instance where the lack of information on the gender of verbs would have resulted in confusion. This permits of the conclusion that as far as machine translation is concerned gender in the Russian verb may well be ignored.

47. THE SYNTHESIS OF RUSSIAN VERB FORMS IN MACHINE TRANSLATION

Z. M. Volotskaya (Moscow)

1. For the synthesis of Russian verb forms in machine translation it is proposed to list in the glossary of stems only the stem of the imperfective aspective of each verb. All the forms of the present, past, and future tense, perfective and imperfective aspect (personal as well as impersonal) are formed from this stem in accordance with definite rules.

2. It is suggested that three types of operation are sufficient to make all possible verbal forms from the single stem: (a) discarding the final letter or letters, (b) adding a letter or letters to the stem on the right, and (c) adding a letter or letters to the stem on the left.

All the individual letters and combinations of letters which are joined to the stem on the left and on the right are assigned by a list and arranged in tables in accordance with a definite system.

3. All the verbs are classified in three groups depending on the method of producing: (a) the forms of the present tense, (b) the forms of the past tense, and (c) the stem of the perfective aspect from the stem of the imperfective aspect.

By class of verbs we mean the total number of verbs that construct a given form in the same way.

4. The information for each verb stem contains the class number of the stem, which indicates the way in which a given form is to be constructed.

48. RUSSIAN SYNTAGMAS

(on the basis of mathematical texts)

Z. M. Volotskaya, Ye. V. Paducheva,
I. N. Shelimova, and A. L. Shumilina
(Moscow)

1. This report discusses the basic types of two-word combinations in subordinate relationship (syntagmas) as found in mathematical texts and by means of which it is possible to construct the rules of formal text analysis (for machine translation).

2. The syntagmas were based on specific word combinations drawn from the texts.

Syntagmas are considered to differ from each other in type of syntactic relations between their component parts. Therefore, not all the morphological and syntactic signs of the words that form the given com-binations served as criteria for relating these combinations to the various syntagmas.

3. A syntagma consists of two components: "governing" and "governed". Each of which is accompanied in the list of syntagmas by certain information.

As a rule, the "syntactic group" is the essential information for the "governing", the "morphological form" for the "governed" component.

4. Words are divided into "syntactic groups" on the basis of the following principle of marking words according to the sign of a common syntactic connection: first, those words which have a single common syntactic connection are separated from the mass of words into one group; then, those words which have another syntactic connection are separated from the same mass, etc. The same words may fall into different groups which consequently appear to be crossing each other.

The separation of syntactic groups not only according to one but according to a combination of signs should lead to a significant increase in the number of syntactic groups and, correspondingly, in the number of syntagmas.

5. The report includes a list of syntagmas, description, and dis-cussion of possible ways of using them in text analysis.

49. SYNTHESIS OF THE RUSSIAN CLAUSE

Z. M. Volotskaya and A. L. Shumilina
(Moscow)

1. Sentence synthesis in machine translation consists of combining words into clauses and clauses into sentences according to the requirements for sentence building in a given language.

2. The aggregate of syntagmas in each sentence that are obtained by analyzing the language from which a translation is made does not constitute an adequate basis for synthesizing sentences of the language into which the translation is made. Correspondences must be established between the languages in question not only on the syntagmatic level but also on the sentence level.

3. A clause is synthesized by inserting a syntagma, i.e., one syntagma as it were overlays and draws into itself another.

4. Each word in the clause of the output language obtains, in addition to the information necessary for translation (number of stem in the output language, number, tense, etc.), the following signs: (a) number of the

syntagma into which it is entering as a governing word (by the first method, cf. below) or as a governed word (by the second method); (b) ordinal numbers of the words (from the input language) with which the given word forms syntagmas.

In combining words into clauses it is more convenient to use the ordinal numbers of words from the sentence of the input language and not the numbers of the output stems because using only the latter might lead to mistakes inasmuch as the sentence may contain several identical lexemes or different ones, but with the same stem.

5. There are two possible ways of synthesizing a clause by means of syntagmas:

(a) Isolating the pivotal syntagmas (predicatives) and successively expanding each component at the expense of the governed words.

(b) Synthesizing a clause by successively combining syntagmas until they are reduced to the predicative. Moreover, each syntagma enters as a single group into a higher rank syntagma as a governed, expanded component.

50. GRAMMATICAL ANALYSIS FOR MACHINE TRANSLATION OF CHINESE INTO RUSSIAN

V. A. Voronin (Moscow)

The system of grammatical analysis for machine translation of Chinese into Russian was based on materials from contemporary scientific and technological texts in mathematics, electrical engineering and construction. It utilized the fundamental works of Soviet and Chinese authors on the modern Chinese language. The system was tested on mathematics articles from the Chinese periodicals Shusyue syuebao (Mathematics Herald) and Shusyue tsin'chzhan' (Successes of the Mathematical Sciences). In constructing the system we did not have the task of solving the extensive and manifold grammatical problems connected with machine translation of literary and socio-political literature. However, we did take cognizance of grammatical phenomena characteristic of Chinese as a whole.

Treatment of the Chinese sentence according to the system of grammatical analysis starts after operation of the glossary and glossary supplement is completed, as a result of which words in the sentence enter the system with concrete relevant meaning and complete lexical characteristics, i.e. with the set of necessary signs.

The special grammatical structure of Chinese possesses an extremely small number of formal means by which one can identify the full morphological properties of the Russian equivalent for the Chinese word within a given lexical unit. Therefore, a Chinese sentence cannot be processed for

machine translation without an analysis of the syntactic structure of the sentence to be translated, which was predetermined by the general principles underlying the system.

The system, operated in the form of routines, consists of two main parts: (1) syntactic analysis of sentences, and (2) production of the morphological characteristics of the Russian equivalent. The entire system includes 9 interrelated, successively functioning routines.

The first part has 4 routines in which the following stages of syntactic analysis are effected in corresponding order:

    (1) Breakdown of the input sentence into simple clauses.

    (2) Separation of attribute+attributed word groups.

    (3) and (4). Separation of other (than attributive) syntactic components of the clause.

The second part of the system has 5 routines of which 4, on the basis of existing syntactic signs, produce the morphological characteristics for the Russian equivalents of all the words in the Chinese sentence. The classes of words mentioned below are handled in the order given:

    (1) Numeral,

    (2) Substantive

    (3) Verb

    (4) Adjective

The operation of the fifth routine consists of changing Chinese word order in accordance with the norms of Russian word order.

The system as a whole comes down to producing the formal signs that reflect in the first part the syntactic function of the word and in the second part the morphological features of the Russian equivalent of the Chinese word.

An adequate, readable translation is ensured by performing a combined lexico--grammatical analysis of the Chinese text put into the machine.

51. APPLICATION OF MACHINE TRANSLATION METHODS TO
THE LEXICAL CODING OF TELEGRAPHIC AND
TELEPHONIC COMMUNICATIONS

V. I. Grigor'yev and G. G. Belonogav (Moscow)

1. Men have been searching from ancient times for the most effective
utilization of the channels of communication. Up to now the main efforts
of engineers and communications experts have been aimed at perfecting the
communication channel proper and at seeking ways of transforming the
signal so as to secure the maximum suitability of the signal to the given
channel. The contents of communications meanwhile remained unchanged.
However, the possibilities have now for the most part become exhausted so
that the problem of findihg means of reducing the size of messages trans-
mitted is becoming increasing urgent.

2. The size of a telegram may be shortened 3-4 times if a lexical
code is used instead of a literal code. A telegraphic communication
that uses lexical coding differs from an ordinary printed letter com-
munication only in that they send not code groups designating letters of
the alphabet, but a code combination designating the ordinal number of
the word according to the dictionary in the memory device plus certain
items containing grammatical information about the word transmitted.

3. The principle of lexical coding of messages has been known
since ancient times. It is employed in various kinds of signal tables, in
the international radio code, and elsewhere. However, in all these cases
coding is done manually, requiring great effort and considerable expendi-
ture of time. The development of computer technology has now made possible
automatization of the process of lexical coding and its wide use in com-
munications.

4. Lexical coding is based on an analysis of the message at the
transmitting end and its subsequent synthesis at the reception end of the
line of communication. This lexical analysis and synthesis of a message
is essentially a simplified form of the analysis and synthesis of a text
produced by machine translation. It is therefore worthwhile, when pre-
paring an algorithm for lexical coding, to make full use of the method of
text analysis and synthesis used for machine translation.

5. Lexical coding has, in addition, several peculiarities. Text
analysis and synthesis in the case of machine translation is aimed at
securing the operation of hieroglyphic conversion--a basic operation in
machine translation. Elimination of hieroglyphic conversion would
lead to considerable simplification of the routines of analysis and
synthesis in the case of lexical coding. On the other hand, with lexical
coding the demand for code economy is pushed to the foreground, whereas it
is of purely secondary significance as far as machine translation is con-
cerned. Lexical coding must rest to a large degree on speech statistics.

In particular, due to the interlinking of analyzer with the devices of the channel of communication, the size of the dictionary cannot be so conveniently large. Available statistics permit limitation of the dictionary of the lexical analyzer to a maximum of 4000 words in ordinary use, which generally make up 97.5% of a literary text. Rare words not found in the dictionary may be transmitted letter-by-letter.

6. Application of the principles of lexical coding to telephonic communication may help greatly in solving the problem of maximum closeness of compression.

52. SOME PROBLEMS IN MACHINE TRANSLATION FROM JAPANESE INTO RUSSIAN

M. B. Yefimov (Moscow)

The purpose of this communication is to set forth some principles involved in analyzing Japanese sentences for machine translation, the principles being characteristic of the Japanese language alone.

A. The primary problem with which we have to deal in analyzing a Japanese sentence is its division into separate words. This is typical chiefly of languages with an ideographic form of script (Japanese, Chinese, etc.). The fact is that words are not separated in a written Japanese text and, consequently, identification of their role in a sentence is quite difficult.

We shall try to show in this report how we made the division in our work.

We began with the fact that the Japanese script uses the signs of a syllabary (kana) along with ideograms.

Thus, the division of a Japanese sentence into separate words breaks down into 3 main steps:

(1) Analysis of portions of sentences containing both ideograms and syllabary.

(2) Analysis of ideographic part.

(3) Analysis of syllabary part.

This operation is closely linked to the operation of the existing Japanese glossary and is, so to speak, one of its parts.

B. Breaking down a sentence into its individual clauses is no less important a problem in Japano-Russian translation and has both practical and theoretical interest.

In this work we are relying chiefly on the rigid structure of the Japanese sentence in which either a verb or a predicate adjective always stands at the end. This enables us infallibly to determine the end of the sentence.

The beginning of the sentence is determined by searching for the subject.

Thus, the entire operation consists of two steps:

    1. Determination of the end of the sentence, and

    2. Determination of the beginning of the sentence.

C. As is true of all languages, the verb constitutes the greatest difficulty in translating from Japanese into Russian.

The strongly developed affixation that is characteristic of Japanese is most clearly marked in the verb.

This determined the cyclical nature of our operation.

We used the fundamental rules of traditional grammar for the analysis of verb endings, relying mainly on the five stems of the Japanese verb.

We have been successful in establishing the necessary grammatical and syntactic criteria for all verbs.

53. WORK ON THE RUSSO-ENGLISH ALGORITHM OF MACHINE TRANSLATION

L. N. Zasorina (Leningrad)

1. Limitation of problems and scope of work. Choice of mathematical text as being most limited in stylistic peculiarities.

Determination of set of programs for Russo-English algorithm. Exclusion of program of differentiating homonyms due to synthetic structure of Russian. Simultaneous work on glossary and morphology program.

2. Combined investigation of short text. Compilation of glossary in which the grammatical form and syntactic relations of the words are registered. Recording of statistical data.

3. Investigation of individual parts of speech, division of words into classes, and preliminary detection of homonymy between the parts of speech.

4. Verb and grammatical information derived from personal forms and nominal forms. Homonomy of participles and adjectives distinguished by taking into account suffixes of full and short forms of participles. Lack of formal--graphic separation of auxiliary and modal verbs from the verb class.

Adjective class comprising adjectives, adverbs in -o, -e, -ski, ordinal numerals, words in the status category. Arrangement in non-specified subclasses. The substantive class including nouns, sub-stantivized words and cardinal numerals (other than odin /one/, dva /two/, tri /three/, chetyre /four/) is distinguished by the abundance of homonymic case forms: intra class homonomy and interclass homonomy. Separation of non-specified subclasses. Triliteral word class. Class of invariable words is characterized by negative separability in the text.

4. Advisability of introducing stem-stripping program. Planning of groups of commands for the individual classes. Many-sided investigation of homonymic coincidences of separable affixes.

5. Problems connected with differentiating grammatical data derived from homonymic affixes. Tables of separable, restrictive lists of letters that precede the separable affixes. Successive separation of affixes from stem (endings and form-constructing suffixes) and storage of grammatical information derived.

Table for verifying matching of preliminary information obtained from affixes and stem glossary.

Method of multistage depositing of grammatical information derived from the glossary and stem-stripping program.

Attempt at dividing grammatical data into two non-crossing fields to reduce the number of tests of possible grammatical forms.

6. Compilation of stem glossary. Determination of general size and limits of glossary. "Lexical article" plan, taking into account input and output information and list of possible forms.

Obtaining pseudostems.

Problems in contracting the glossary by separating word-building suffixes and prefixes.

7. General routine for processing words: stem-stripping program, stem glossary, morphology program. Obtaining input information for the syntactic program.

54. WORK ON A HINDUSTANI (HINDI) - RUSSIAN ALGORITHM OF MACHINE TRANSLATION

T. Ye. Katenina (Leningrad)

1. The development of a Hindi-Russian algorithm is very important for similar work in the field of Indian languages--both Indo-Aryan and Dravidian. The structure of Hindustani is in the main analytical, although the traces of ancient inflection and agglutinative elements--a new synthesis--play a definite role. The scientific style of Hindi prose is characterized by a more or less definite word order close to that of the Dravidian languages. Numerous phrases containing non-conjugated verb forms, equivalent to subordinate clauses, constitute the main difficulty for machine translation. Scientific texts are characterized by an abundance of Sanskritisms which are frequently translated loan words of international (European) terms.

2. Hindi writing, phonetic for the most part, is therefore especially convenient for an electric reading device. To record texts we worked out a mechanical transcription based on the Russian alphabet without complicated signs and diacritics. In addition statistics justified our combining several Hindustani sounds.

3. The set of programs for machine translation is as follows: (1) glossary of stems (2) morphology program (3) postposition program (4) syntactic program (5) program for differentiating homonyms (6) list of idioms (7) a translation program of compound words may be required for some kinds of scientific texts.

4. In order to avoid superfluous information we adopted the following hypothesis for the syntactic analysis of a simple sentence: (1) the first noun substantive in a direct or active case is the subject (2) the verb in the last place in the sentence is the predicate (3) if the verb is not a copula, the noun substantive in the next-to-last place in the sentence with the postposition ko or in the direct case (not the subject) is the direct object.

We have determined the necessary minimum of morphological information-- but which requires statistical confirmation in individual cases--to be: (1) for the noun substantive--number case (direct, active, indirect), (2) for nominal adjective--number (may be important to determine the number of noun substantives with zero ending of direct case plural number), (3) for the verb -- tense, mood: number (to determine the number of the same noun substantives); voice. A check of the text showed that the overwhelming majority of simple sentences as well as the constituent parts of complex sentences may be analyzed in accordance with these rules.

6. Among the basic problems requiring a solution for subsequent
work in constructing a Hindustani-Russian algorithm are: (1) elucidation
of rules for analyzing complex sentences and equivalent phrases with non-
conjugated verb forms, (2) clarification on a statistical basis of the
need to design a program analyzing compound words that would be compulsory
for all kinds of texts.

55. AN ALGORITHM FOR TRANSLATING ENGLISH TEXTS ON RADIO ENGINEERING INTO RUSSIAN

K. V. Komissarova (Gorki)

The translation rules and glossary have been worked out with regard
for the characteristics of English texts dealing with radio engineering.

The translation process is divided into 2 main parts: analysis of
English sentences and synthesis of Russian sentences.

Analysis of an English text is based on a syntactic analysis of the
sentence. The grammatical function of a word is determined by morphological
and syntactic analysis according to rules grouped by the parts of speech.

The glossary contains more than 500 words in general use and specialized
technical terms.

56. AUTOMATIZATION OF TRANSLATION PROGRAMMING

O. S. Kulagina (Moscow)

1. Long, tedious process of constructing translation programs causes
need to automatize programming. Requirements of translation programs and
impossibility of using existing programming programs. Formulation of
problem of automatizing translation programming.

2. Breakdown of translation algorithms into operators. Types of
operators and functions of each. Parameters of operators.

3. Preparation of translation algorithm for translation: recording
of algorithm in the form of sequence of simple rules, transition from this
recording to operator, automatic construction of translation program ac-
cording to operator recording of algorithm by means of compiling program.

4. Compiling program, its structure. Some features of structure of
programs obtained by the method described.

57. A FRENCH-RUSSIAN TRANSLATION ALGORITHM

O. S. Kulagina (Moscow)

(1) Formulation of problem: translation of mathematical texts. De-
mands for quality in translation, cases requiring editing.

(2) Structure of glossary for machine translation, features. Glossary
information and purpose. Glossary of phrases.

(3) Principles in constructing translation algorithm. Structure of
algorithm and order of operation. Word look-up in glossary. Treatment
of phrases. Differentiation of homonyms and analysis of polysemants, order
of operation of rules for differentiating homonyms. Analysis of French
sentence, problems. Sequence of handling parts of speech during analysis.
Character of information obtained through analysis. Change of word order
in translation. Synthesis of Russian sentence; order of operation of
synthesizing rules and how they differ from analyzing rules.

(4) Supplementing and correcting algorithm on the basis of experimental
translations (greater precision in rules for differentiating homonyms,
change in handling of adjectives, separation of morphological from syntactic
analysis).

58. DETERMINATION OF SYNTACTIC CONNECTIONS FOR FORMULAS
IN RUSSIAN MATHEMATICAL TEXTS

M. M. Langleben (Moscow)

1. We call "formulas" all text elements not found in a mechanical
glossary in processing a text (surname, mathematical formulas, foreign
references, neologisms, etc.). "Formulas", like words to be translated,
require the ascertaining of syntactic connections in the text to be
analyzed, i.e., the identification of formulas that form part of one of
the previously given syntagmas.

2. The analysis of a "formula" is broken down into 2 parts:

   (A) testing the formula proper for the presence of any word-
        changing suffixes, the sequence of tests being determined by
        frequency of the cases.

   (B) analysis of its environment (words and punctuation marks).
        This begins only after all the "formulas" contained in the
        given segment of text have passed through part A.

3. The following order of ascertaining the possible syntactic
connections for "formulas" is advisable in that it eliminates the possi-
bility of establishing false syntagmas:

(a) the formula acts as an adjective for a substantive standing on the right;

(b) the formula is a name with a substantive standing on the left;

(c) the formula forms part of a prepositional phrase;

(d) the formula forms part of a syntagma with an adjective requiring the dative case (RAVNYI /equal7, KRATNYI /multiple7);

(e) the formula forms part of a syntagma with an adjective in the comparative degree replacing a substantive in the genitive case;

(f) the formula replaces a governing substantive in an "adjective + substantive" syntagma;

(g) the formula acts as a predicative combination.

These last are used to check various syntagmas with a verb; the function of a formula with a verb is chiefly determined by its position on the right or left of the formula, not by the form of the verb.

4. Since the analysis of "formulas" is a basic part of the routine developed for the language as a whole, it will be performed piecemeal at various stages of the total analysis.

59. ELIMINATION OF MORPHOLOGICAL AND SYNTACTIC HOMONOMY IN ANALYZING ENGLISH TEXTS

M. M. Langleben and Ye. V. Paducheva (Moscow)

1. Those words in a dictionary of stems that cannot be identified as a fixed part of speech, i.e. "attempt" (verb, noun), "cool" (adjective, verb), and "further" (adverb, adjective), etc., are handled as follows:

If a word can be a noun and a verb or an adjective and a verb, it is inserted in a dictionary of substantives or verbs, respectively. Those word-changing suffixes that can readily identify one part of speech to which a word belongs (-ed, -ing, but not -s) are listed in a table of word-building suffixes, i.e. if the word has one of these endings, the part of speech will be revealed after morphological analysis. However, homonymic stems do not require any changes in the analysis routine provided for the other words. (This method is based on a suggestion by A. I. Smirnitskii who defined conversion as word building by means of paradigms).

2. If the part of speech cannot be readily determined by morphological analysis of (zero ending in word stems) "They attempt", "The attempt", homonymic ending--"he attempts", "The attempts--or the parts of speech which have no word-changing forms are homonymic--"further" (adjective, adverb)--

the word is assigned several syntactic functions corresponding to the possibilities of the word to enter a syntagma as a substantive, verb, etc. The possible functions are examined in a definite order and a syntagma is established for the given word, depending on whether certain words are present in the sentence; thereafter all the remaining functions listed are dropped out except that for which the syntagma was found.

3. Similarly, homonomy in -ing forms, -ed forms, etc. is eliminated by successive tests for the presence of certain syntagmas in the sentence.

60. THE SUPERFLUOUSNESS OF RUSSIAN ADJECTIVE INFLECTION

N. N. Leont'yeva and G. N. Vavilova (Moscow)

1. In machine translation from Russian the procedures for handling the inflection of adjectives are quite cumbersome. The machine has to perform a double task: first, to investigate the inflection of the adjective, then to search for the substantive with which the adjective agrees.

There is an easier way of relating an adjective to the substantive with which it agrees, a way that ignores inflection in most cases.

2. When a Russian text is analyzed, it usually turns out that adjective inflection is superfluous as far as translation in concerned. It merely indicates the agreement of the given adjective with a certain substantive.

3. An adjective may be related to the substantive with which it agrees without analysis of its inflection by using the adjective's position in the sentence.

An adjective - attributive most frequently occupies with respect to the substantive with which it agrees a definite position: it stands either before this substantive or after it, following a comma.

Accordingly, it is possible to formulate two rough rules for relating an adjective to its substantive:

(a) Relate the adjective to the nearest substantive on the right;

(b) If there is no substantive on the right, relate the adjective to a substantive that is followed by a comma.

4. However, relating an adjective to a substantive in accordance with these rules alone may turn out to be incorrect.

Therefore, a number of individual tests must be performed before finally deciding the problem of relating an adjective: is the adjective part of the nominal constituent of the predicate, is it included in a formula, does it govern the following noun with or without a preposition (VYZVEDENNYI IZ FORMULY /deduced from the formula/, RAVNYI NULYU /equal to zero/).

5. After these checks the machine either relates the adjective to the substantive without regard to its inflection or, if it cannot dispense with it, analyzes the inflection of the adjective.

6. An analysis of mathematical texts shows that without investigation of inflection it is possible to relate more than 85% of all adjectives to the appropriate substantives. The remaining 10-15% of the adjectives requires an analysis of the inflections.

7. In calculating the number of adjectives we excluded short adjectives, the relative KOTORYI /which/, cases where the adjective is part of a formula, cases of ellipsis (the adjective is present, but not the noun with which it agrees, e.g. OTLICHAYETSYA OT RASSMOTRENNYKH B ETOM PARAGRAFE /it differs from the (things) considered in this paragraph/.

8. The practicability of a method to ascertain the possibility of ignoring adjective inflection has still not been proved. This will require further work on texts as well as more experience with machine translation, taking cognizance of technical difficulties.

Nevertheless, the suggested routine for relating an adjective to its substantive by position criteria will retain its value, even if the necessity for investigating the inflections of all adjectives is demonstrated, since inflection is merely one of the factors that control the correct relating of an adjective to its substantive by position criteria.

61. AN ALGORITHM OF MACHINE TRANSLATION FROM ENGLISH INTO RUSSIAN

T. N. Moloshnaya (Moscow)

I. (1) Different possibilities for formalizing linguistic data in different languages.

(2) Advantages of a structural-syntactic analysis of English.

II. (1) Classification of English and Russian words according to formal criteria.

(2) Grammatical configurations constructed from isolated classes of words.

III. Analysis of English sentence structure according to grammatical configurations.

(1) Replacement of grammatical configuration by its chief member.

(2) Sequence of ascertaining grammatical configurations in the sentence to be analyzed.

IV. Synthesis of Russian sentence structure according to grammatical configurations.

(1) Substitution of the English grammatical configuration used by the corresponding Russian configuration.

(2) Morphological formation of Russian sentence structure.

(3) Definition of grammatical forms of words in the Russian sentence.

V. Elimination of lexico-grammatical homonomy in the English sentence on the basis of:

(1) morphological data,

(2) syntactic data

VI. Tests of machine analysis of English sentence structure.

62. A DEVICE FOR THE READING OF ORDINARY PRINTED MATERIAL BY THE BLIND

R. S. Muratov (Sverdlovsk)

1. Conversion of the graphic form of letters in a printed text into electrical signals is achieved by breaking down the group of photosensitive elements as they move along the line of text.

2. Electrical impulses generated when photosensitive elements are blacked out switch on electronic relays which, in turn, switch on a tactile or phonic signalling instrument.

3. The form of the signals (of successive formation of elementary signals corresponding to each zone of disintegration) expresses the graphic peculiarities of the letters and other marks in the text.

4. Correct reading of the signals requires preliminary instruction by a reader.

63. ANALYSIS OF PUNCTUATION MARKS DURING MACHINE TRANSLATION FROM RUSSIAN

T. N. Nikolayeva (Moscow)

1. The purpose of this operation is to obtain the distinguishing features of punctuation marks during machine translation.

2. In translation from Russian each word in the sentence must receive definite morphological and syntactic signs. The required signs are obtained in different ways for each part of speech. In particular, in order to determine the case and number of substantives it is necessary to know the correlative position of the parts of speech within the limits of the closed sentence /ZAMKNUTOVO PREDLOZHENIYA/. However, most Russian sentences are complicated by parenthetical and set-off /i.e. by commas-OBOSOBLENNYM/ constructions, subordinate clauses, etc.

Hence, to obtain the precise grammatical signs it is necessary to break down a complex sentence into simpler components, dividing the main clause from the subordinate clauses and separating the set-off and parenthetical phrases.

Thus, the final goal of the analysis of punctuation marks is to:

(a) separate simple clauses from the body of the complex sentence, to find the boundaries of the simple clause within the sentence;

(b) separate similar members of the clause;

(c) help the subsequent elucidation of interrelations between the individual parts of the punctuated complex sentence;

(d) determine a group of similar members.

2. /sic/ The analysis is made within a single complex sentence. Accordingly, "simple" and "multipurpose" punctuation marks are distinguished. The simple ones (period, exclamation point, question mark, and dots) serve as the boundaries of a complex sentence.

Multipurpose marks (comma, dash, and colon) unite simple clauses into a complex clause, introduce subordinate clauses, and separate parenthetical and set-off constructions.

We are devoting the bulk of our attention to the multipurpose marks. In a clause they may serve, according to Prof. A. B. Shapiro's terminology, to "divide" or to "separate". We are also paying special attention to the problem of distinguishing between single and non-single punctuation marks (e.g. those used at the end of a set-off phrase and the beginning of a subordinate clause etc.).

- 73 -

3. As a result of the analysis, all the multipurpose punctuation marks receive one of the following signs:

     (1) parenthetical (i.e. separating parenthetical words and phrases);

     (2) setting off (separating participial and verbal-adverb phrases as well as set-off attributives and appositives);

     (3) similar-simple (dividing similar members of a sentence);

     (4) similar-complex (demarcating the parts of a compound sentence);

     (5) dissimilar (i.e. introducing a subordinate clause).

4. Separation of the simple clauses occurs within the limits of the complex whole according to our data.

The entire process of analyzing punctuation marks can be divided into 3 stages:

     (1) Separation of the purely parenthetical constructions takes place in the analysis glossary where the words that may be used parenthetically or that are a basic part of a parenthetical phrase undergo special analysis, after which the punctuation marks that separate them receive an appropriate information sign.

     (2) Processing of punctuation marks by the "Punctuation Marks" routine, where the basic analysis of all the punctuation marks takes place.

     (3) Breakdown of the sentence into its constituent parts—separation of parenthetical and set-off constructions, dividing of simple clauses, etc. Here the occurrence of a "non-single" mark is extremely important. This routine also provides for insertion of a sentence-demarcating punctuation mark where necessary.

5. The "Analysis of Punctuation Marks" routine consists of several parts, each of which corresponds roughly to a given punctuation mark.

Within each part several checks are made on a number of individual factors that determine the function of the multipurpose punctuation marks.

These factors include the presence of verbs with the sign "LF" (LICHNAYA FORMA) /personal form/ on both sides of a given mark (or on one side of it), the presence of verbs with the sign "NELICHNAYA FORMA" /non-personal form/, the place of a substantive with the sign FS ("FORMA SLOVARNAYA") /dictionary form/ in respect to the given mark, the separation of words belonging to a given lexical group, etc.

6. As a result of our investigation, all the punctuation marks are provided with the requisite distinguishing features and the analysis is performed accordingly within the separated simple units.

64. SOME PROBLEMS CONNECTED WITH THE ANALYSIS OF COMPLEX SENTENCES AND CLAUSES WITH SIMILAR MEMBERS

Ye. V. Paducheva (Moscow)

1. The following problems must be solved in connection with the syntactic analysis of complex sentences and clauses with similar members:

(a) To distinguish between a syntagma with similar members and clause coordination (the difficulties in solving this problem are explained by the fact that most of the co-ordinating conjunctions (L.,ILI, NO) /and, or, but/ may connect both similar members of clauses and entire clauses and therefore they cannot serve as a trustworthy sign either of clause boundary or of syntagma with co-ordinating connective /SOCHINITEL'NOI SVAZ'YU/);

(b) To separate words interlinked by a co-ordinating connective, having divided them from the words governed by them.

2. For this purpose we propose the following method of analyzing sentences with co-ordinating conjunctions (only 2-member combinations are considered for the time being): The sentence is cut up into "chunks", the limits of which are co-ordinating conjunctions, and the syntactic analysis is performed within the chunks; if after completion of syntactic analysis within the chunk no words remain without a governor, it means that the conjunction connects two clauses; if, however, such words remain, it means that the sentence contains similar members. Words lacking a governor are, for the most part, members of a co-ordinating syntagma.

3. When words are combined into a coordinating syntagma, the concept of "sameness of form" /RAVNOOFORMLENNOSTI/ is used. "Sameness of form" is the coincidence of several of their morphological and syntactic signs. The same form is sought beyond the chunk for a word that lacks a governor within the chunk and a coordinating syntagma is thereby established. (This must be refined somewhat due to the possible absence of agreement in number for words with the chunk, etc.).

4. This method of analysis is feasible for Russian because a word normally contains all the information regarding the possible syntactic connections for it (with some exceptions,—compare, e.g., the homonomy of cases, which may make the syntactic function of a word in the chunk indefinite). This method is impracticable in English (e.g. the syntactic functions of a substantive are determined wholly by its position

after a transitive verb, before another substantive, etc.; therefore, superfluous "subjects" would appear after the division into chunks is made.

However, some of the difficulties mentioned for Russian disappear in English during the analysis of a sentence with co-ordinating conjunctions due to the rigid word order and preferential position of the governed word after the governor. English syntagmas with co-ordinating connectives are determined at the same time as the others during the course of syntactic analysis.

5. Some methods of fixing the boundaries of a simple clause inside a complex clause are indicated.

66. MACHINE TRANSLATION OF COMPOUND NOUNS FROM GERMAN INTO RUSSIAN

V. V. Parshin (Moscow)

1. The extensive use of compounds in German, particularly in scientific and technical literature, has made it necessary to work out universal rules for their translation.

Formulation of such rules makes possible a significant reduction in the size of the dictionary and the translation of compounds, provided that the components are known.

Universal rules for the translation of compounds are deduced from a structural-semantic analysis of the constituent words. Determination of semantic connections between them ensures an adequate translation.

The author's investigations do not pretend to be a complete and final solution to the problem of translating compounds. They are merely an initial, empirical attempt at working out the basic principles and methods that would permit of a more or less successful translation at the first stage.

2. The existence of the following types of connections between the stems of compounds has been demonstrated by an analysis of concrete linguistic material (individual original works on mathematics and a German-Russian polytechnical dictionary):

1. Relation of the sum to the constituents,

2. Relation of a part to the whole,

3. Object or subject of an action to the action,

4. Object of the bearer of a quality to the quality,

5. Object of a determiner to the thing determined.

Translation of the first component of compound words, the internal connections of whose components relate to the first four types, is effected by producing a Russian equivalent in the genitive case.

If the last type of connection is present, the first component is translated in two ways: by a adjective and the production of a Russian equivalent in the genitive case.

Polysemia causes a certain type of connection for each meaning of the word. Therefore, a semantic analysis of the components is necessary to differentiate the types of relations between the constituent elements.

Differentiating the relations of a part to the whole and the relation of a determiner to the thing determined is the most difficult of all.

3. A special case is the translation of compounds consisting of three components. It is important here to establish the co-subordination of determining stems to the determined stem, which is done by subjecting them to analysis in pairs.

Three-component words are translated in accordance with the rules for translating two-stem words.

4. Compounds of the input text are broken down into constituent stems by the superposition of stems included in the dictionary, taking into account connecting consonants and rejected endings.

5. The principles and methods of translating German compounds into Russian, as set forth above, can serve as the basis for a definitive, detailed solution of one of the most complicated lexicographical problems in German.

66. PROPER NOUNS IN MACHINE TRANSLATION

A. V. Superanskaya (Moscow)

1. Proper nouns are unavoidably present in every scientific test.

2. In the present state of development the machine translates a text, but leaves proper nouns just the way they are, printing them in Latin letters.

3. Since the number of proper nouns increases as one proceeds from selective to continuous translation, the question of the desirability of automatizing the process of transcribing proper nouns arises.

- 77 -

4. Proper nouns are not always written, read, and pronounced in all languages in accordance with the rules for common nouns.

5. Proper nouns are international. The same nouns are encountered among peoples of different nationality. People move from country to country and publish their papers in different countries in different languages. That is the reason for the difficulty in determining the nationality of a noun and, accordingly, the rules by which it should be transcribed.

6. There is much inconsistency in the current transcription of nouns. The need to unify the transcription and eliminate the lack of uniformity is long overdue.

7. Due to the limitless memory potentialities of the machine and the difficulty of mechanical analytical transcription, it is more efficient to store proper nouns as a whole in the machine's memory. Consequently, if it encountered such a noun in a text, the machine would locate it in the glossary and deliver the answer (simple or in several variants, depending on the linguistic origin of the noun and on existing traditions). This would help to make transcription uniform; and it could be accompanied by a printed glossary to match.

67. WORK ON A BURMESE-RUSSIAN ALGORITHM OF MACHINE TRANSLATION

O. A. Timofeyeva (Leningrad)

1. The syllabic nature of Burmese writing requires the elaboration of a special program by which an electrical reading device can handle a Burmese text.

2. We are compelled to restrict the algorithm to the literary form of Burmese speech owing to the sharp divergences between the written and contemporary spoken languages.

3. A highly developed word-building root structure that crosses with a form-building root structure makes it necessary to have a special word-building program, the purpose of which is to separate lexical from morphological phenomena.

4. The development of agglutination and the rudiments of internal inflection require the construction of a complicated morphological program for handling the abundant and varied grammatical information contained in the Burmese word.

5. The absence of a rigid order for nominal members of the Burmese sentence complicates the syntactic program, which cannot be effected without the preliminary operation of the morphological program.

68. WORK ON AN ARABIC-RUSSIAN ALGORITHM OF
MACHINE TRANSLATION

O. B. Frolova (Leningrad)

I. Items from newspapers are used as texts in machine translation from
Arabic to Russian.

II. The main principles in working on an Arabic-Russian algorithm of
machine translation, as contrasted with those of traditional grammar,
are as follows:

(a) Only the written form of the language with the infixes consonants
and long vowels is considered, whereas all the existing grammars take into
account the short vowels, which are not normally noted in writing. For
Arabic two algorithms, differing in principle, are necessary: one for the
spoken language, the other for the written; the two variants are not re-
ducible to each other.

(b) The traditional dictionary arranged by roots is replaced by a
dictionary arranged by stems.

(c) For convenience in transliterating Arabic letters into Russian
letters, the latter are used with no additional signs of any kind.

III. The programs making up the algorithm are as follows: (1) stem-
stripping (2) address (3) morphological (4) syntactic (5) dictionary of
stems (6) table of prepositions (7) glossary of idioms and phrases (8)
program for distinguishing homonyms.

IV. Work on the stem-stripping program:

(a) Initial variations of this program provided for cutting off the
stems, prefixes, and suffixes; the glossary increased considerably, how-
ever, due to pseudostems.

(b) An important factor in simplifying this program was the idea of
a reject /OTKAZNOI/ glossary which was later developed into the idea of an
address   used in other algorithms too.

(c) The stem-stripping program includes the following rules:

(1) Out of the 28 letters of the Arabic alphabet 10 letters may
be joined as non-radicals to the beginning of a word: these
are certain conjunctions and prepositions, the definite
article, and verbal prefixes.

- 79 -

(2) In the case of words that do not contain initial non-radical letters, it is necessary to refer at once to the address; endings and suffixes are automatically stripped upon comparing the words with the stems found in the address.

(3) Some of these initial non-radical letters, which when cut off reveal an insignificant number of pseudostems, are first transferred to the end of the words and converted into suffixes that are kept apart; the words are then sought in the address.

(4) Words with remaining initial non-radical letters, which if cut off would result in a large number of pseudostems, are first checked in the address; if they are not found there, the non-radicals are transferred to the end of the words, and the words are again looked up in the address. Checking for their presence in the address is not equivalent to extracting from the address all the information relating to the stem.

69. EXPERIMENTAL TRANSLATIONS FROM FRENCH INTO RUSSIAN

G. V. Chekova (Moscow)

Devising of algorithms for translation from French to Russian.

Sequence of operations for translation programs. Changes in programs and coding of glossary on the basis of experimental translations produced by the machine.

Utilization of scales in translation programs.

Programming characteristics; scope of programs and glossary; operations utilized in translation programs; numerical characteristics of translation programs.

Basic demands made of a special translation machine.

Examples of translations produced by the STRELA machine in 1957-1958.

70. ESTABLISHMENT OF SYNTACTIC CUES FOR
PREPOSITIONAL PHRASES

I. N. Shelimova (Moscow)

1. The object in making a syntactic analysis of prepositional phrases consisting of either a preposition and substantive standing to the right of it or a preposition and pronoun immediately adjacent to it on the right is

- 80 -

to include these prepositional phrase in syntagmas. It is necessary, therefore, to find a word in the sentence with which the prepositional phrase forms a syntagma.

2. There are no complications in drawing up the rules for the formal analysis of prepositional phrases if a word that belongs to a class of words capable of forming a syntagma with the prepositional phrase is found immediately to the left of the prepositional phrase. The only exception is a case where a noun stands next to the prepositional phrase. Thus, if there is any verbal form - infinitive, participle (short or full), verbal adverb, or adjective (full or short) - or special group of invariable words on the left of the prepositional phrase, the prepositional phrase forms a syntagma with this particular word.

3. If on the left of the prepositional phrase is a word that belongs to a class of words with which the prepositional phrase does not generally form a syntagma (pronouns, adverbs, particles, conjunctions) or the prepositional phrase stands at the very beginning of the sentence, then the word with which the prepositional phrase forms a syntagma must be searched for in the following order:

(a) Search to the left for the next word with which the prepositional phrase may become a syntagma, excluding a noun, i.e. search for any form of verb, adjective or special kind of invariable word. A prepositional phrase may unite in a syntagma with several of the classes of words listed after it fulfills a series of conditions.

(b) Search to the right for the next word belonging to the class of verbs (except the full participle and verbal adverb) or a word from the special group of invariable words or a short adjective. Actually while searching for a word on the right, with which the prepositional phrase may form a syntagma, we are looking for a word in the predicate of the sentence.

4. If a prepositional phrase stands next to a noun (immediately to the left of the noun), the rule for establishing the syntagma constituted by this phrase is not general for prepositional phrases with different prepositions.

5. Therefore, any of the following may be significant in determining the rules for analyzing prepositional phrases with a number of prepositions:

(a) The lexical composition of the prepositional phrase itself;

(b) Does the prepositional phrase have on its left a noun which by virtue of its syntactic or lexical properties is such that its connection with the prepositional phrase must be regarded as certain?

(c) Does the sentence have any verbal form that by virtue of syntactic or lexical properties must be regarded as necessarily connected with a given prepositional phrase?

- 81 -

6. The structure of the sentence is particularly important in establishing the rules of syntactic analysis for prepositional phrases with several other prepositions (e.g. v /in/ with the prepositional case and dlya /for/). In order to determine the regular syntactic cues for the prepositional phrases mentioned, it is necessary in certain cases to know if the prepositional phrase stands before or after the predicate or which syntagma contains the noun that is followed by the prepositional phrase. Sometimes it is important to know whether or not this noun in turn forms a prepositional phrase with certain prepositions (e.g. v rezul'tate /as a result of/, posle /after/, etc.) because in such a case a prepositional phrase with dlya or v cannot be related to this noun.

71. CORRELATION BETWEEN 3RD PERSON PERSONAL PRONOUNS AND THE NOUNS FOR WHICH THEY SUBSTITUTE

A. L. Shumilina (Moscow)

1. In machine translation the 3rd person personal pronouns of one language cannot be mechanically substituted for the corresponding pronouns of another language since gender is not an inherent sign of every pronoun, but depends on the gender of the corresponding noun, which is accidental as far as they are concerned and specific for the different languages.

2. The following formal data must be obtained first if the correlation between a pronoun and the corresponding substantive is to be established:

(a) The boundaries of the clauses (no cognizance is taken of the differences between the boundaries of clauses within sentences and sentence boundaries);

(b) The grammatical properties of the substantives and 3rd person personal pronouns (gender, number, case);

(c) The syntactic relations and specific syntactic functions of the substantives;

(d) The order of substantives in the clause;

(e) Certain sequences of syntactically related words (e.g. expanded attributes).

3. A substantive for which a given pronoun is used must "correspond grammatically" to this pronoun. By grammatical correspondence we mean the correspondence between substantive and pronoun in number (correspondence in number will in several cases differ from the conventional) and gender (in the singular).

- 82 -

4. The way to determine the corresponding ("unknown") substantive is, for the most part, as follows:

The search for the grammatically corresponding word is made only to the left of the given pronoun (omitting the previously determined elements in the clause).

A. Within a zero $^{(1)}$ clause ⌐(1)Clauses subject to analysis are numbered: zero (0) ≡ a clause within which the given pronoun is found, first (1) ≡ next clause to the left of the zero, second (2) = next clause to the left of the first, etc.⌐

   (a) For pronouns in the nominative case, the only possible unknown substantive may be one with a sign of the "grammatical subject" (this concept is defined beforehand).

   (b) For pronouns in other than the nominative case, the unknown word is the substantive that is closest to the given pronoun, but with certain restrictions (e.g. the unknown substantive must not form a single word combination with the given pronoun, nor must it be the middle word or word on the extreme right in a chain of genitive cases, if the word on the extreme left satisfies the sign of "grammatical correspondence", etc.)

B. Within the first, second...nth clause (The analysis is made successively within the 1st 2nd...nth clause until the word that satisfies our requirements is found).

For pronouns both in the nominative and in other cases, a word with a sign of the "grammatical subject" is considered first; in the event that there is no grammatical correspondence between the pronoun and the "grammatical subject" found, we pass on to a word with a sign of the "grammatical direct object", then to the substantive that is closest to the right boundary of the 1st or nth clause (taking into account the various restrictions already determined).

5. Similar work in the future may, with appropriate additions (animateness in nouns and other criteria), be significant from the point of view of practical stylistics, i.e. it may create the possibility of determining certain purely formal rules for using 3rd person personal pronouns on the basis of the laws of the language itself.

* * *
⬝ E N D ⬝

US JPRS/DC
DUPONT 7-4240